National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

## NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

## AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

NL-339 (r. 88/04) c

Canada

UNIVERSITY OF ALBERTA

# ACOUSTIC CHARACTERISTICS OF PATHOLOGIC LARYNGES: AN ACOUSTICAL AND PERCEPTUAL STUDY OF CHANGES IN DYSPHONIA WITH VOICE THERAPY.

BY

KATHERINE LISA WARD

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND
RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE
IN SPEECH-LANGUAGE PATHOLOGY

DEPARTMENT OF SPEECH PATHOLOGY AND AUDIOLOGY

EDMONTON, ALBERTA

Spring, 1992

National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-73244-X

Canada

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Katherine Lisa Ward

TITLE OF THESIS: Acoustic characteristics of pathologic larynges: An acoustical and perceptual study of changes in dysphonia with voice therapy.

DEGREE: Master of Science in Speech-Language Pathology

YEAR THIS DEGREE GRANTED: 1992

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

Katherine Ward

10807 65 Avenue, Edmonton, AB T6H 1V8
Student's permanent address

Date: April 15 1992

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **ACOUSTIC CHARACTERISTICS OF PATHOLOGIC LARYNGES: AN ACOUSTICAL AND PERCEPTUAL STUDY OF CHANGES IN DYSPHONIA WITH VOICE THERAPY** submitted by Katherine Lisa Ward in partial fulfillment of the requirements for the degree of Master of Science, Department of Speech-Language Pathology and Audiology.

Anne Putnam Rochet
Dr. Anne Herbert Rochet

Darlene Broda Masiuk
Ms. Darlene Broda-Masiuk

Dr. Anton Rozsypal

Dr. Frank Wilson

Date:

# ABSTRACT

The relationship between acoustical analysis and perceptual evaluation methods for monitoring changes in dysphonia was investigated via a single-subject A-B time series design. Tape-recorded samples of sustained /a/ from four adult females with vocal nodules were digitized (20 kHz) and measured for four acoustical parameters: jitter, shimmer, fundamental frequency and signal-to-noise ratio (SNR), using CSpeech3 (Milenkovic, 1989). Graphic analysis and visual inspection were used to document changes in time series data for the acoustical variables across baseline and treatment periods. Pearson product-moment correlations were used to explore the relation among the average difference values for the acoustical variables for the vowel /a/ from the last week of baseline and the last week of therapy. Three speech-language pathologists rated before- and after-treatment voice samples in a paired-comparison paradigm on a visual analog scale (VAS) of voice quality. The relationship between the acoustical and perceptual data was investigated by correlating the jitter magnitude difference scores (in ms) between end-of-baseline and end-of-treatment paired samples of sustained /a/ with the perceived difference in relative ratings of dysphonia (in mm on VAS) for the same paired samples.

Results showed that measures of jitter, shimmer, and SNR in the midportions of sustained productions of /a/ did not appear to change significantly or in a consistent direction with voice treatment. Fundamental frequency values showed more absolute change with treatment and tended to change in a manner more consistent with the change in each subject's vocal nodules. Of all possible correlations among the acoustical variables, only three between jitter and shimmer were significant. All judges were able to perceive a difference in severity of dysphonia between end-of-baseline and end-of-treatment paired samples of sustained /a/. Jitter levels did not correlate significantly with perceptual ratings of voice however. Interpretation of these results must be tempered by the recognition that these findings have not been replicated, the sampled number of data points was small and a millimetre scale was arbitrarily applied to the VAS to provide data for correlations and for the purpose of discussion.

# ACKNOWLEDGEMENTS

I am grateful to Dr. Rochet for her supervision and encouragement throughout this project. Completion of this thesis would not have been possible without her generous support.

I also wish to express my sincere thanks to the other members of my committee for reading this manuscript and supporting my efforts throughout this project. I appreciate the information Dr. Rozsypal provided, particularly with respect to the acoustical analysis procedures. I wish to thank Ms. Darlene Broda-Masiuk for her helpful suggestions regarding voice treatment. As well, I am grateful to Dr. Wilson for performing the Flexible Fiberoptic Nasendoscopic examinations necessary for the initiation and completion of this research.

I would also like to express my deepest gratitude to Chris Upton, my fiancé. His continuous support, encouragement and sense of humour helped me to see the light at the end of the tunnel.

# TABLE OF CONTENTS

Chapter Page

I. INTRODUCTION . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1

II. LITERATURE REVIEW. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

  Pitch extraction procedure . . . . . . . . . . . . . . . . . . . . . . . 3

  Temporal resolution . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

  Sample duration. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

  Perturbation formulae . . . . . . . . . . . . . . . . . . . . . . . . . . 7

    Absolute measures . . . . . . . . . . . . . . . . . . . . . . 8

    Fundamental frequency related measures. . . . . . . . 9

    Short-window measures. . . . . . . . . . . . . . . . . . . 10

  Variables influencing jitter magnitude . . . . . . . . . . . . . . . 13

    Voice onset and termination. . . . . . . . . . . . . . . . 13

    Vowels . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14

    Sex. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

    Menstrual cycle. . . . . . . . . . . . . . . . . . . . . . . . . 17

    Age. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19

    Fundamental frequency. . . . . . . . . . . . . . . . . . . 21

  Studies differentiating normals from pathological groups. . 23

  Perceptual studies. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28

III. METHODOLOGY. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39

  Design . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39

    Experimental variables. . . . . . . . . . . . . . . . . . . . 39

    Experimental phases . . . . . . . . . . . . . . . . . . . . . 39

  Subjects. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41

Chapter Page

Materials/equipment. . . . . . . . . . . . . . . . . . . . . . . . 42

Procedures. . . . . . . . . . . . . . . . . . . . . . . . . . . . 43

Subject selection. . . . . . . . . . . . . . . . . . . . . . . . 43

Voice data recordings. . . . . . . . . . . . . . . . . . . . . . 44

Voice treatment . . . . . . . . . . . . . . . . . . . . . . . . 46

Acoustic data acquisition and analysis . . . . . . . . . 48

Analog-to-digital conversion and analysis. . . . . . . 48

Perceptual data acquisition . . . . . . . . . . . . . . . . 50

Reliability . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52

Acoustical data. . . . . . . . . . . . . . . . . . . . . . . . 52

Perceptual data. . . . . . . . . . . . . . . . . . . . . . . . 53

IV. RESULTS. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55

Flexible fiberoptic video nasendoscopy (FFVN). . . . . . . . 55

Menstrual cycle. . . . . . . . . . . . . . . . . . . . . . . . . . . 56

Acoustical parameters. . . . . . . . . . . . . . . . . . . . . . . 56

Jitter . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 56

Shimmer . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 60

Fundamental Frequency ($F_0$). . . . . . . . . . . . . . . 64

Signal-to-noise ratio (SNR). . . . . . . . . . . . . . . . . 67

Multiple correlations. . . . . . . . . . . . . . . . . . . . . . . 70

Perceptual ratings. . . . . . . . . . . . . . . . . . . . . . . . . 74

Reliability. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 74

Correlations with jitter. . . . . . . . . . . . . . . . . . . . . . 76

V. DISCUSSION. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80

Menstrual cycle . . . . . . . . . . . . . . . . . . . . . . . . . . 80

Chapter Page

  Baseline stability . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 82
  Comparing the results of different studies. . . . . . . . . . . . 83
  Jitter.. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 84
  Shimmer. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 86
  Fundamental frequency. . . . . . . . . . . . . . . . . . . . . . . . 88
  SNR. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 90
  Multiple correlations. . . . . . . . . . . . . . . . . . . . . . . . . . . 92
  Limitations and suggestions for future research . . . . . . . 101
  Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 104
References. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 106
Appendices. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 116

# LIST OF TABLES

Tables Page

1. Subject information. . . . . 42

2. Intrajudge reliability. . . . . 53

3. Interjudge reliability. . . . . 54

4. Summary of each subject's post-treatment vocal nodule status and acoustical data. . . . . 56

5. Summary of average jitter, shimmer, SNR and $F_0$ values within and between phases. . . . . 71

6. Multiple correlations among average jitter, shimmer, Fo and SNR difference values. . . . . 72

7. Perceptual difference ratings correlated with jitter difference magnitudes. . 77

8. Summary of end-of-baseline and end-of-treatment jitter difference values paired with judges' perceptual difference ratings. . . . . 97

# LIST OF FIGURES

Figures Page

1. Instrumental array for voice data acquisition. . . . . . . . . . . . . . . . . . . . . 45
2. Instrumental array for digital analysis of the acoustical data. . . . . . . . . . . 50
3. Instrumental array for perceptual data acquisition. . . . . . . . . . . . . . . . . . 51
4. Visual analogue scale for perceptual ratings of dysphonia. . . . . . . . . . . . 52
5. Average session jitter across baseline and treatment phases. . . . . . . . . . . 57
6. Average session shimmer across baseline and treatment phases. . . . . . . . 62
7. Average session $F_0$ across baseline and treatment phases. . . . . . . . . . . . 65
8. Average session SNR across baseline and treatment phases . . . . . . . . . . 68
9. Scattergrams for multiple correlations among the acoustical variables . . . . 73
10. Perceptual ratings of end-of-treatment voice samples. . . . . . . . . . . . . . . 75
11. Scattergrams for each judge's perceived difference score against jitter difference values. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 78

# LIST OF APPENDICES

Appendix Page

A. Subject's consent form allowing the primary investigator to contact subjects by phone. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 116

B. Telephone screening form . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 117

C. Vocal nodules treatment study information . . . . . . . . . . . . . . . . . . . . . 119

D. Subject's participation consent form. . . . . . . . . . . . . . . . . . . . . . . . 122

E. Voice case history. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 124

F. Voice treatment information for subjects . . . . . . . . . . . . . . . . . . . . . . 127

G. Instructions for perceptual ratings. . . . . . . . . . . . . . . . . . . . . . . . . . 132

# CHAPTER I. INTRODUCTION

Acoustic analysis is a non-invasive procedure used by speech-language pathologists to help document pathological changes in the vocal mechanism and improvements in the voice associated with therapy. It may also be utilized as a potential screening device for early detection of laryngeal disease (Ramig, Scherer, Titze & Ringel, 1988). Two parameters measured by this method are jitter and shimmer. Small variations in the speech waveform period (jitter) and amplitude (shimmer) are known to be natural characteristics of normal voice (Lieberman, 1961; 1963) and are necessary for the natural quality of synthetic speech (Holmes, 1962). These variations are probably related to the periodicity of the neuro-muscular phonatory control system and have been used to explain hoarse vocal quality in the speech signal. Researchers have demonstrated that both normal speakers and those with vocal pathologies display frequency and amplitude perturbations (Moore & Thompson, 1965; Moore & von Leden, 1958). Speakers with vocal pathologies demonstrated greater perturbation values than normal speakers, however (Deal & Emanuel, 1978; Iwata & Von Leden, 1970). These increased frequency and/or amplitude perturbations were also associated with hoarseness and correlated positively with severity of laryngeal pathology and perceptual roughness ratings (Coleman, 1969; Deal & Emanuel, 1978). The use of objective acoustic parameters could lead to more effective communication among speech-language pathologists and be more reliable than clinical (i.e., perceptual) judgements alone. Acoustic parameters could also provide more precise measurements of change in voice characteristics before and after therapy and thus indicate the relative efficacy within and among different treatment protocols. Significant correlations between the acoustical perturbation values and perceptual evaluations before and after therapy would enhance the applicability of the acoustic measures.

A number of studies have measured differences in jitter between normal and pathological voices. However, results have been somewhat conflicting, primarily due to differences in measurement methods and varying controls for factors known to affect jitter, such as speaker sex, age, fundamental frequency, intensity, vowel differences and intonation patterns. For example, most of the research has been done on men's voices; few studies have included women and children among their subject groups, and additional studies controlling for gender differences and other pertinent variables are needed to establish normative values for both sexes and age groups and examine jitter magnitudes for normal voice versus those associated with specific types of laryngeal pathologies.

This research project was designed to explore the relationship between acoustical and perceptual voice analysis methods with the focus on jitter, because there is a need to document changes in jitter values associated with therapy for specific laryngeal pathologies and how these values relate to perceptual judgements by clinicians. This research also attempted to control for some of the factors that are thought to affect jitter and other acoustic variables but previously have been uncontrolled in studies of this kind. Specifically, jitter and shimmer magnitudes, fundamental frequency ($F_0$) and signal-to-noise ratio (SNR) were measured in the sustained vowel /a/ produced by four adult females with vocal nodules, before, and during voice therapy. The relationship between these acoustical measures and listeners' perceptions of dysphonia in the subjects' voices was also investigated.

# CHAPTER II. LITERATURE REVIEW

Several methodological issues must be considered to determine jitter levels accurately and to compare the results among investigations. These issues include the pitch extraction procedures used in marking the individual (fundamental) periods of vocal fold movement, the sample rate and duration, the formulae for the calculation of jitter and a number of other variables known to influence jitter. The following review outlines each of these issues. As well it will discuss studies which differentiate normal from pathological voices, studies which correlate jitter values with perceptual measures, and relationships between the results of such studies and the results of this study.

## Pitch extraction procedure

Hundreds of pitch determination methods and algorithms have been developed. These may be subdivided into two main groups; the first group includes short-term analysis methods that deal with the signal in a more universal way, and the second group includes time-domain methods that process the signal period-by-period. Short-term analysis methods measure the average pitch period duration. Voice perturbation measurement requires time-domain pitch determination algorithms (PDA's) (Crystal & Jackson, 1970; Horii 1975, 1979) or a combination of a non-stationary short-term analysis PDA and a time-domain PDA (Moser & Kittel, 1977).

The time-domain strategies used most often in jitter studies are 1) peak-picking, which counts the number of positive peaks occurring each second, 2) zero-crossing detection, which counts the number of times the wave crosses the zero line in one direction or the other, per second, and 3) inverse filtering, which cancels the vocal tract transfer function and the glottal shaping function to estimate the period of the voice source. Milenkovic (1987) and Horii (1979) state that the peak-picking methods of pitch extraction are limited by their need for a very high speech signal sampling rate (> 40kHz) to obtain the temporal resolution required to measure jitter and shimmer. Processing the speech signal at a high sampling rate is expensive, and not all computer programs allow processing of the signal in this manner. Peak-picking sometimes requires manual intervention to "consistently identify the waveform peak between successive pitch periods in many natural waveforms" (Milenkovic, 1987, p. 529). As an alternative, Milenkovic (1987) employs the short-term autocorrelation function for pitch tracking. In this method, measures of jitter and shimmer are based on the averages of complete pitch periods of the speech waveform, thus eliminating the "reliance on a single waveform peak" (Milenkovic, 1987, p. 530). A moving observation interval is applied to the speech waveform so that the perturbation measures derived from the autocorrelation function can be updated on a sample-by-sample basis. This allows for an initial estimate of the fundamental period and automatic tracking of its variation. Interpolating samples of the autocorrelation function allows for a high degree of temporal resolution in jitter analysis. Deem, Manning, Knack and Matesich (1989) suggest that interpolation for the estimation of jitter provides an advantage, particularly if low sampling rates are used. In addition, Titze, Horii and Scherer (1987) state that interpolation is necessary if sampling is done at less than 500 samples per cycle. This is equivalent to sampling rates of 50 and 100 kHz for adult male and female voices, respectively. Parabolic interpolation of the samples of the autocorrelation function was employed in the present research.

## Temporal Resolution

In the past, techniques used to obtain frequency perturbation measures have relied on measurement of high-speed motion pictures of the glottal waveforms (Lieberman, 1963; Moore & Thompson, 1965), optical oscillographic records obtained from laryngeal contact microphone recordings (Beckett, 1969; Koike, 1973), or high-speed motion pictures of the acoustic phonatory signal displayed on an oscilloscope (Hecker & Kreul, 1971; Hollien, Michel & Doherty, 1973; Murry & Doherty, 1980). More recent studies have used computer analysis of digitally stored acoustic waveforms (Horii, 1979, 1980, 1982; Wilcox & Horii, 1980; Karnell, Scherer & Fischer, 1991).

Differences in the technique for capturing and measuring the acoustic signal become critical only when they affect the temporal resolution of measurement. In view of the small magnitudes of jitter in the vocal signal, this is the most critical factor in perturbation studies. Time resolutions reported in the literature range from 500 μs (Moore & Thompson, 1965) with hand measurements from a phonelograph trace, to 2 μs (Ludlow, Bassich, Connor, Coulter & Lee, 1987) with special purpose software to directly process the signal. The most frequently used resolutions are near the high-resolution end of this range, 50 or 25 μs.

Horii (1979) demonstrated the importance of the effects of temporal resolution on vocal jitter measurements. Three 5-second recorded voice samples from one subject with average frequencies of 150, 210, and 260 Hz were quantized and analyzed three times at each of 5 different sampling rates (5, 10, 20, 40, and 80 kHz). The 80 and 40 kHz sampling conditions yielded comparable jitter magnitude results except at the highest $F_0$ level. In contrast, the 10 and 5 kHz sampling conditions produced larger mean jitter magnitudes than the other sampling conditions, with the 5 kHz condition yielding the largest amount of jitter. However, the mean jitter was not always over-estimated by poorer temporal resolutions. The 20 kHz sampling condition for the low frequency voice and the 40 kHz sampling condition for the high frequency voice signal under-estimated the mean jitter in those instances. Examination of data for each sampling condition also showed greater jitter magnitude variability as the sampling rate decreased. Horii's results suggest that the relatively large jitter values that have been reported for normal sustained phonation, such as mean jitter values greater than 0.02ms (Zemlin, 1962, cited by Horii, 1979; Moore & Thompson, 1965, cited by Horii, 1979), probably were inflated due to the particular temporal resolution employed. Horii cautions that the results of his study "need to be interpreted with limitations imposed by the 0.025 ms temporal resolution in mind" (1979, p.13).

### Sample duration

Another methodological consideration in the calculation of jitter and shimmer magnitudes concerns the number of cycles that are required for accurate measurement of these acoustical features in sustained vowels. Titze, Horii and Scherer (1987) investigated the effect of the number of cycles and tokens included for jitter and shimmer analysis. A single subject with no signs of vocal pathology was required to sustain 32 tokens of the vowel /a/ for 7 seconds each at a comfortable pitch and loudness. The tokens were digitized at 20 kHz, and interpolation of the samples was employed. Single windows of varying cycle length (10, 20, 40, 80 and 160 cycles) were selected for analysis within each vowel token. Results of this study suggested that 20 to 30 cycles was a sufficient window length for jitter and shimmer analysis. Titze et al. cautioned that "this result is not guaranteed, however, for pathologic voices" (p.259). Deem, Manning, Knack and Matesich (1989) explored this issue further by measuring jitter in 5 male and 5 female subjects for each of three vowels (/i/, /a/, /u/) across five sample durations (10, 20, 40, 80 and 160 cycles). Recorded samples were digitized at 100 kHz and analyzed via a microcomputer program (PAEJ). Results indicated that sampling only 10 cycles resulted in lower mean jitter values and greater variability in jitter across gender and vowel. Greater stability of jitter measures was found in longer samples, particularly those including 40 or more cycles; analysis of 80 and 160 cycles resulted in similar jitter values. Similar to the results of Titze at al.

(1987), the results of Deeming et al.(1989) show that sample durations greater than 40 cycles were not necessary for speakers with normal voices.

The length of the sample duration required for measurement of jitter in speakers with pathological voices may be longer than that required for speakers with normal voices. Karnell (1991) examined jitter and shimmer magnitude changes in 18 subjects referred for voice evaluation. Subjects were asked to produce the vowel /a/, 3 times for approximately 5 seconds. Results of this study showed that as many as 190 cycles may be required before jitter asymptotes, and as many as 130 cycles before shimmer asymptotes. However, 110 cycles is probably adequate for both jitter and shimmer measures. Karnell (1991) proposes that procedural issues, such as establishing window analysis length, signal recording conditions, computer analysis procedures and perturbation algorithms must be specified in reports of vocal analysis research before data from different laboratories can be considered for comparison.

## Perturbation Formulae

A variety of different measures has been proposed for quantification of speech waveform pitch perturbations. These measures continue to be refined in an attempt to distinguish normal jitter values from those that characterize pathological voices and to be sensitive to improvements in or deterioration of a person's vocal function. Three general categories of measures are apparent in the literature: 1) absolute measures, which ignore the speaker's fundamental frequency and determine the percentage of perturbations equal to or greater than an absolute value in milliseconds, or the percentage of adjacent period differences that differ in algebraic sign, 2) frequency (relative) related measures, which relate the perturbation values to the average fundamental frequency; and 3) short-window measures, which determine perturbations in terms of their distance from the average perturbation over a small number of periods. Jitter and shimmer measures may be expressed using various units. These include percent jitter (%, derived from the percentage of perturbations equal to or greater than an absolute, or the percentage of perturbations between adjacent periods that differ in sign), or mean jitter in milliseconds (ms) or microseconds (μs) (derived from one of the algorithms presented in the next section). Shimmer may be expressed in percent (%, derived from the percentage of perturbations equal to or greater than an absolute, or the percentage of perturbations between adjacent amplitudes that differ in sign), or mean shimmer in decibels (derived from one of the algorithms suggested by Baken, 1987).

Absolute measures. Lieberman (1961, 1963) devised the first measure of jitter known as the "Pitch Perturbation Factor" (PPF). A pitch period perturbation refers to the time difference between the duration of successive periods in the voice signal. Lieberman (1963) hypothesized that $F_0$ perturbations reflect changes in 1) glottal periodicity, 2) alterations of the glottal waveform, and 3) variations of vocal tract configuration that result in phase shifts of the acoustic wave. He defined this parameter as "the percentage of all perturbations equal to or greater than 0.5 milliseconds in a vocalic segment of connected speech" (Lieberman, 1963, p. 344). Lieberman demonstrated that pathological voices generally have larger perturbation factors than normal voices with comparable $F_0$'s and that the perturbation factor is sensitive to the size and location of tissue growths in the larynx. Later investigators also demonstrated that the pitch perturbation factor discriminates between phonation produced by normal and pathological larynges (Crystal & Jackson, 1970; Iwata & von Leden, 1970; Murry & Doherty, 1980).

Hecker and Kreul (1971) presented another absolute perturbation measure, the "Directional Perturbation Factor" (DPF). The DPF refers to "the percentage of the total number of (period time) differences for which there is a change in algebraic sign"

(Hecker & Kreul, 1971, p. 1279). In contrast to Lieberman's (1961) measure, it ignores the magnitude of period perturbation and is concerned solely with the number of times the frequency change shifts direction (Baken, 1987). Hecker and Kreul (1971) found that the DPF distinguished jitter values for a group of 5 normal speakers from jitter values for 5 speakers with cancer of the vocal folds. Lieberman's PPF did not separate the two groups.

Fundamental Frequency related measures. The magnitude of frequency perturbation is not independent of mean fundamental frequency ($F_0$). There is a significant correlation between these parameters, with larger jitter magnitudes associated with longer fundamental periods, or low pitches (Lieberman, 1963; Becket, 1969; Koike, 1973; Horii, 1979, 1980), and smaller perturbation magnitudes associated with shorter periods, or high pitches. This presents a potential problem when screening individuals with vocal pathologies who also have higher $F_0$'s; by virtue of the high $F_0$, they may exhibit vocal perturbations falling within the normal range. While there is no way to normalize exactly for the effect of mean $F_0$ on jitter magnitude (Horii, 1979; Hollien et al., 1973) and thereby achieve a jitter index uninfluenced by fundamental frequency, the best compromise is to generate a ratio of some form of mean perturbation to mean period (Baken, 1987). Two measures that purport to meet this need are Jitter Ratio (Horii, 1979) and Jitter Factor (Hollien et al., 1973). The measure is called the Jitter Ratio (JR) when calculated in terms of period durations:

$$\text{Jitter ratio} = \frac{\frac{1}{n-1}\left[\sum_{i=1}^{n-1} \left| P_i - P_{i+1} \right| \right]}{\frac{1}{n}\sum_{i=1}^{n} P_i} \times 1000$$

where $P_i$ = period of the $i^{th}$ cycle, in ms, and n = number of periods in the sample.

Jitter Factor is the frequency-based equivalent of Jitter Ratio (Hollien et al., 1973). The average differences between the frequencies of adjacent cycles are divided by the average overall frequency, and multiplied by 100.

$$\text{Jitter factor} = \frac{\frac{1}{n-1}\left[\sum_{i=1}^{n-1} \left| F_i - F_{i+1} \right| \right]}{\frac{1}{n}\sum_{i=1}^{n} F_i} \times 100$$

where n = number of periods in the sample, and F = fundamental frequency of the $i^{th}$ cycle in Hz.

Short-Window Measures. Relatively slow and steady increases and decreases in pitch, as well as abrupt, rapid, quasi-random frequency shifts characterize changes in vocal fundamental frequency during conversation. The former are related to linguistic variables (such as prosodic features of stress and intonation), and hence are volitional and superimposed on vocal output as linguistic content demands. The latter are sudden, presumably involuntary changes and should be reflected in jitter magnitude levels

(Baken, 1987). In running speech, the two types of fundamental frequency modulation occur simultaneously. Any measure that evaluates one must inevitably include the effects of the other on acoustic measurements. Even during sustained vowels, $F_0$ may undergo slow changes that could inflate the measure of jitter (Baken, 1987). This potential problem is not addressed by the perturbation formulae discussed previously. The "relative average perturbation" (RAP) (Koike, 1973), also known as the "frequency perturbation quotient" (Takahashi & Koike, 1975), was presented as a solution to account for artifacts of voice/speech production that may influence jitter. This formula uses a form of straight-line-averaging across three fundamental periods that reduces the effect of relatively slow changes in the $F_0$ on jitter calculations.

In the "Relative Average Perturbation " measure (RAP):

$$\mathrm{RAP} = \frac{\frac{1}{n-2}\sum_{i=2}^{n-1}\left|\frac{P_{i-1}+P_i+P_{i+1}}{} - P_i\right|}{\frac{1}{n}\sum_{i=1}^{n} P_i}$$

where P(i),i= 1,2......,N, denote the successive pitch periods, and n = the total number of periods.

The numerator is composed of the average difference between actual periods and their three-point estimates, and the denominator is composed of the average period, which is included to compensate for the change in mean absolute jitter that occurs with change in $F_0$ (Baken, 1987). Koike (1973) was able to distinguish the voices of 30 subjects with laryngeal disease from those of 30 speakers with normal larynges using the RAP.

Kitajima, Tanabe, and Isshiki (1975) presented another variation on this averaging procedure. They defined a fundamental frequency perturbation as the average of 100 successive frequency perturbations from an all-voiced phrase. A frequency perturbation was defined as the difference between a measured frequency and a five-point weighted average (in a least-squares sense) centered around the measured value. Results of their study indicated that normal female voices exhibit larger $F_0$ perturbation values than normal male voices, which would be expected since the adult female fundamental frequency is generally higher than that of the adult male. Takahashi and Koike (1975) combined Koike's earlier results into two time-domain measures of the pathological voice; the Frequency Perturbation Quotient (FPQ) and the Amplitude Perturbation Quotient (APQ). The FPQ is analogous to Koike's RAP, except that the instantaneous $F_0$ (defined as the reciprocal of the pitch period) is substituted for the pitch period. An eleven-point, rather than a three-point, averaging window is used. Davis (1979) defined a pitch period perturbation quotient (PPQ) analogous to Koike's RAP. However rather than using fixed three-point or eleven-point averaging windows, Davis systematically varied the window size and found that five-point averaging windows produce the best perturbation measures for a reliable normal/pathological voice discrimination. He also found that perturbation measures based on the residue signal are better for discriminating among voices than those based on the direct speech signal. Residue signals refer to indirect signals derived from inverse filtering techniques. Direct signals refer to those obtained from throat contact microphones and microphone-transduced radiated sound pressures. Inverse filtering cancels the effects of the vocal tract and lips on the speech signal to approximate the voiced source signal, thus avoiding "supraglottal affects whose resonance characteristics are presumably unaffected by laryngeal pathology" (Davis, 1979, p.275).

Milenkovic (1987) employs another semi-averaging procedure for calculating perturbation values. This procedure was used in the present investigation. Milenkovic's method uses the short-term autocorrelation function for pitch tracking and determines measures of jitter and shimmer based on the average of entire pitch periods of the speech waveform. Perturbation measures are continually updated by applying a moving observation interval (≤ pitch period) to the speech waveform. Parabolic interpolation is applied to improve the temporal resolution of the signal. Milenkovic (1987) demonstrated the accuracy of the autocorrelation method in measuring low levels of perturbation in both synthetic and natural voice waveforms using a medium sampling rate (8.3 kHz).

## Variables influencing jitter magnitude

In addition to measurement methods, a number of source and vocal tract filter variables are known to influence jitter magnitude. These include voice onset and offset characteristics, type of vowel, speaker sex, age, and fundamental frequency. All are factors to be considered when comparing results from different perturbation studies.

Voice onset and termination. Greater frequency perturbation is noted at the beginning and end of sustained vowels than during their steady-state midsections (Lieberman, 1961; Horii, 1979). Koike (1973) found that in men and women with normal voices, the first 17 glottal cycles of a soft initiation (start of quiet voicing) had a mean relative average perturbation of 0.0276 in comparison to 0.0046 ms for steady-state phonation. The "most appropriate and most frequently used phonatory task appears to be the midportion of a single sustained vowel produced at a natural, comfortable pitch and loudness level when changes in perturbation caused by anatomic-physiologic conditions of the larynx are in question" (Zyski, Bull, McDonald & Johns, 1984, p.

191). Baken (1987) suggests that clinicians evaluate sustained vowels no less than a half-second after voice initiation, unless characteristics of voice onset, *per se*, are being examined.

Vowels. The jitter values reported by investigators for different vowels have differed. Wilcox and Horii (1980) were among the first to note jitter and/or shimmer differences among vowels. They found that /u/ was associated with significantly smaller jitter (0.55%) than /a/ or /i/ (0.68% and >0.69%, respectively). Horii (1980) also found both jitter and shimmer magnitudes to be smallest for /u/ , intermediate for /i/ and greatest for /a/. In contrast, Sorensen and Horii (1983) found relatively low jitter values for /a/ (0.71%), intermediate for /u/ (0.86%) and high for /i/ (0.96%). Johnson and Michel (1969) also reported a trend toward greater jitter magnitude for high vowels than low vowels. Horii (1982) measured jitter and shimmer from the mid 3 seconds of eight English vowels /I, i, ae, a, o, u, u and ʌ/ each sustained for at least 5 seconds by twenty normal subjects. The vocal signal was transduced by means of a miniature accelerometer placed on the throat. Jitter and shimmer magnitudes were calculated at a sampling rate of 40kHz via a computerized program (SEARP) that uses a peak-picking method to find individual periods. Statistical analysis revealed no significant differences in either percent jitter or shimmer magnitudes among the eight vowels. An average jitter value of 0.75% and an average shimmer value of 0.17dB were obtained across the eight vowels and 20 subjects. These jitter values were in general agreement with data reported for other normal voices. Jacob (1968), Hollien et al. (1973) and Horii (1979) reported jitter values of about 0.5-1.0% for phonations produced at comfortable intensity and frequency levels. Horii (1982) controlled for soft initiation-termination, and phonations sustained for a comfortable length of time and intensity. These are important variables to consider according to Ramig (1980); if subjects are expected "to phonate at a specific intensity, and/or as long as possible, jitter /shimmer values are expected to increase" (Ramig (1980) as paraphrased by Horii (1982, p. 14). Caution is therefore necessary in comparing "normative" data or performance of studies that have used various analysis methods for calculating jitter and shimmer. Results of this study indicate that 1) when the accelerometer was used as a laryngeal-signal transducer, the shimmer values were considerably smaller than those measured in the airborne signal transduced by a microphone, and 2) there may be little systematic difference in jitter or shimmer among sustained phonation of various vowels. Milenkovic (1987) measured jitter and shimmer from the voice tape recordings of three male and three females sustaining the vowels /a/, /u/ and /i/ at a comfortable frequency and intensity. Considerable variation in jitter and shimmer was found among the vowels. He reported that the differences observed in jitter and shimmer values among vowels may be due to the "difference in vocal tract impulse response from one vowel to the next rather than changes in the underlying voice source." Further he suggested that "it may be important to control more strictly for vowel formant frequencies in making jitter and shimmer comparisons between subject groups"(Milenkovic, 1987 p. 537).

The issue of jitter magnitude associated with specific vowels remains unresolved. Differences could be due to different data collection methods, the perturbation formulae used for analysis and varying controls for subject variables such as age, sex, and speaking conditions.

Sex. Few studies have addressed sex differences for jitter. Although many researchers have included both males and females in their subject groups, results have most often been pooled between genders for comparisons between healthy and pathological voices. This has clouded the issue of whether or not actual differences in jitter values exist between the sexes. The few studies of vocal jitter that have separated subjects according to gender provide equivocal results.

Sorensen and Horii (1983) investigated vocal jitter and shimmer during the sustained phonation of /a/, /i/ and /u/ by 20 adult females using the SEARP automatic peak-picking analysis program. Subjects were asked to sustain productions of the vowels for 5 seconds at a comfortable frequency and intensity level. The middle three seconds of each vowel were digitized at 40 kHz and stored on computer tape. Results indicated that jitter values were low for /a/ (0.71%), high for /i/ (0.96%) and intermediate for /u/ (0.86%). Overall average jitter was 0.84%. Jitter values of 1% or less have been reported for the middle portions of sustained vowels produced by male speakers (Jacob, 1968; Koike, 1973; Hollien et al. 1973; Horii, 1979). The values in the Sorenson and Horii (1983) study of female voices are within this range but are greater than the same values reported by Horii (1980) for males. He reported values of 0.61%, 0.60% and 0.72%, respectively, for men's productions of /a/, /i/ and /u/. Using an analysis of variance for their data, Sorensen and Horii (1983) found both a main effect for sex, with larger overall magnitudes for females, and a sex-by-vowel interaction. T-tests revealed a significant difference between the males and females for /i/ and /u/, but not for /a/ ($p< .05$).

These investigations were followed by those of Hartmann and von Cramon (1984) and Haji, Horiguchi, Baer and Gould (1986). Hartmann and von Cramon (1984) measured jitter in the vowels /a, e, i, o, u/ sustained by 10 males and 7 females with normal voices. Their voiced signals were digitized at 20 kHz, and the durations of individual periods for jitter analysis were determined using the autocorrelation method. Results revealed that female voices exhibited significantly larger jitter magnitudes than those of males ($p< .005$) when the 5 vowel values were averaged together for each speaker. The effects for vowels were not explored. In contrast, Haji et al. (1986)

reported no sex effect for jitter measured during sustained /a/ in 18 males and 12 females. These investigators also employed a sampling rate of 20 kHz, but jitter was measured from electroglottograph (EGG) signals. The individual periods were marked from base-line crossings, and jitter was reported in semitones. Similarly, Orlikoff and Baken (1990) reported that jitter values of six men (23-47 years) and six women (21-44 years) did not seem to be significantly different when values were averaged over samples representing a significant portion of their phonational frequency range. These authors measured absolute and relative jitter in the vowel /a/ sustained at each of 6 semitone levels. Jitter was measured from EGG signals at a sampling rate of 100 kHz.

The issue of sex effects on jitter remains to be resolved. Differences in results may depend on the vowels measured, the differences in the methods used to transduce, measure and analyze the signals, or all of these factors.

### Menstrual cycle

Several researchers have attempted to relate acoustic changes in the female voice to particular stages of the menstrual cycle. This interest began with the suggestion by Frable (1962) that just prior to menses a drop in estrogen and progesterone levels resulted in mass increases of the vocal folds due to edema. With the resulting increase in mass one would expect a decrease in $F_0$ prior to menses. Studies examining this relationship have yielded conflicting results. Several researchers have reported no significant differences between $F_0$ at premenstruation compared with other stages of the menstrual cycle in women with and without vocal training (Brown & Hollien, 1981, 1982; Wilson & Purvis, 1980; Silverman & Zimmer, 1978). Wilson and Purvis (1980) explored the variations of speaking fundamental frequency (SFF) in trained singers. Speaking fundamental frequency was measured from Visicorder traces of one phrase of a standard reading passage and /a/ sung at subjects highest and lowest $F_0$'s during ovulation, premenstruation and menses. No significant changes in SFF were noted during the three stages of the cycle.

Silverman and Zimmer (1978) investigated the effects of the menstrual cycle on vowel spectral noise levels in sustained productions of /a/, /i/ and /u/ produced by 20 students with no vocal training. They found no significant differences between vowel spectral noise levels gathered at premenstruation and ovulation. These authors were able to replicate this result in another sample of 27 undergraduate students. Brown and Rothman (1985) measured harmonics-to-noise ratio (HNR) and vocal frequency perturbation across the menstrual cycle in teachers and singers. They hypothesized that vocal fold mass changes associated with premenstruation would result in increased hoarseness. No significant differences were found in HNR or jitter from recordings obtained a day within the premenstrual period and a day outside the premenstrual period. However, they reported significantly reduced jitter and increased HNR during menstruation when compared with the premenstrual and nonpremenstrual days.

Higgins and Saxman (1989) measured jitter in the sustained vowels /i/ and /u/ transduced by a throat contact microphone and analyzed via a computer. Recordings were collected every other day over a 33-day sampling period for 10 women (aged 19 - 25 years) and in 5 men (aged 19-26 years) who served as a control group. Perturbation values for the vowels /i/ and /u/ were combined as no significant differences between them for women were found. Results for the women were analyzed with respect to ovarian hormone fluctuations inferred from basal body temperature data and subjects' reports of the onset of menstruation. Mean jitter values for each subject were computed and plotted over time. No similarity in the pattern of variation over time was seen between groups or subjects. They found no significant differences in the average jitter magnitudes of women at premenstruation and the onset of menstruation. A change greater than 1 SD from the subject's mean perturbation value was considered to be significant. They did however discover large upward and downward changes in jitter magnitudes at the time of ovulation for most subjects. Specifically when standardized jitter factor values from ovulation and postovulation were pooled to test the significance of jitter around the time of ovulation, change in jitter around the time of ovulation was significantly greater than that which would have been predicted based on chance alone. Higgins and Saxman (1989) speculate that findings of increased variation in jitter at ovulation lends support to the idea that when ovarian hormones fluctuate the vibratory stability of the vocal folds is affected. Further they state that these influences may be strong enough to influence the voices of non-singers during conditions of normal speaking.

Age. There is a growing body of literature concerning the general effects of aging on body physiology and changes in acoustical variables of voice associated with age-related changes in the larynx. Perceived changes in the aging voice may be due to physical changes in the laryngeal mechanism that are a result of the normal aging process (Benjamin, 1981). Kahane notes the following changes: "ossification and calcification of the laryngeal cartilages, loss of tissue elasticity, thinning of mucosa, reduced secretory function and degeneration and atrophy of laryngeal muscles" (1981, pp. 24-26). These age-induced changes in the larynx may introduce structural instability and cause increased jitter magnitude (Kahane, 1981; 1987). Investigators examining the effect of aging on changes in vocal perturbation are concerned with 1) distinguishing perturbation increases due to the normal aging process from those caused by vocal pathologies, and 2) determining separate normative values for different age ranges.

Wilcox and Horii (1980) examined jitter in the sustained vowels /i, a and u/ produced by 20 young adults (18.4-25.8 yrs) and 20 older adults (60.5-80.0 yrs). The older adults showed more jitter (0.060 ms) than the young group (0.046 ms), with mean jitter ratios of 0.73 and 0.55 ms, respectively, although there were large individual differences and overlap among the young and old subjects in their study. The jitter ratio values ranged from 0.41 to 0.84 ms in the younger group and from 0.44 to 1.37 ms in the older group. Thus, although the older speakers as a group displayed a wider range in jitter, many of their younger counterparts fell within the older group's range. This seems to suggest that chronological age itself is not a good predictor of jitter values. Benjamin (1981), using a Honeywell Visicorder to produce an oscillogram for $F_0$ perturbation analysis, also found a greater percentage of frequency perturbation in the sustained vowel productions of older male and female speakers, aged 68-82 years (10.26%) than in those produced by younger counterparts, aged 21-35 years (2.66%). Heiberger and Horii (1982) found significantly more jitter and shimmer values for the vowels /a, u, and i/ produced by a group of older males (60 to 80 years) than for the same vowels sustained by a group of younger males.

Ramig and Ringel (1983) measured jitter and shimmer magnitude in vowels produced for a comfortable duration by 48 men representing three age groups (averaging 31, 53 and 68 years). In contrast to Heiberger and Horii's (1982) study, they found no significant differences among the three groups. Ramig and Ringel (1983) hypothesized that because physiological aging of the body may not directly reflect chronological aging in all persons, the large intersubject variability within chronological age groups could be partially attributed to physiological differences among the members of each group. The forty-eight males representing three chronological age groups (25-35, 45-55, and 65-75 years) in their study were divided into two groups (good and poor physical condition) on the basis of a number of physiological indices (resting heart rate, resting systolic and diastolic blood pressure, percent body fat and forced vital capacity). Results showed that subjects in good physical condition produced maximum sustained-vowel durations with significantly less percent jitter and shimmer magnitudes than those produced by subjects of similar chronological age but in poorer condition. The age-by-physical condition interactions did not reach statistical significance, but the authors noted that difference trends were most apparent among the elderly speakers. On the basis of these results, they suggested that measures of age-related changes in body physiology, in addition to chronological age, should be used to classify normative data on acoustical measures of voice.

Orlikoff (1990) measured mean fundamental frequency, mean jitter and shimmer (absolute and relative), as well as the standard deviation of $F_0$, SPL, and peak-to-peak vocal amplitude in the sustained productions of /a/ produced by eighteen adult males. Subjects were divided equally into groups of healthy young and elderly men and elderly men diagnosed with chronic atherosclerosis. Significant differences were found between the healthy young and healthy elderly subjects on measures of $F_0$ and amplitude standard deviation, percent jitter, and shimmer. When the younger subjects were compared to the elderly atherosclerosis group these differences in the acoustic variables were magnified. Percent jitter was the only acoustic variable that differentiated the two elderly groups; the atherosclerosis group's sustained phonations were associated with increases in intra- and intersubject short- and long-term variability compared to those produced by the healthy elderly men. Orlikoff (1990) suggests that elderly speakers " may be more prone to vocal disruption in the face of pathology" (p. 450).

### Fundamental frequency.

Jitter has been investigated in both connected speech and sustained vowel phonation. Caution is necessary in interpreting jitter measures derived from connected speech, "since both systematic perturbation due to phonetic context, stress and intonation, and random perturbation associated with mechano-physiologic limitations of the glottal source coexist in such voice signals" (Horii, 1979, pp. 5-6). Observations concerning the relationship between $F_0$ level and $F_0$ perturbations are equivocal (Horii, 1979). As mentioned previously, many researchers have employed jitter ratios, rather than absolute jitter, because of the dependence of absolute jitter size on $F_0$ level.

Jacob (1968) reported that the largest jitter ratios were associated with sustained vowels produced at low pitch and low vocal intensity. Beckett (1969) and Jacob (1968) found that jitter ratios decreased as a function of increasing $F_0$. In contrast, Montgomery (1967) did not observe any systematic relation between magnitude of jitter ratio and $F_0$, and Hollien et al. (1973) reported the tendency for the $F_0$ jitter ratio to increase at high $F_0$. Horii (1979) measured jitter in the middle segments of the sustained vowel /i/ produced by six male adults, 28-43 years of age. The vowel was produced at eleven fundamental frequencies ranging from 98 to 298 Hz at levels between 70-80 dB SPL. Mean jitter ranged from 24-52 µsec across the eleven frequency levels, revealing a tendency for the mean jitter to increase as $F_0$ decreased.

In a recent study, Orlikoff and Baken (1990) further explored the relationship between $F_0$ and mean jitter for normal adult men and women. Six men (aged 23-47 years) and six women (aged 21-44 years) were asked to produce maximal prolongations of the vowel /a/ at each of six semitone levels. Absolute and relative jitter values were computed from an EGG and microcomputer. Regression analysis used to explore the linear regression of absolute jitter magnitude on $F_0$, revealed a significant ($p < 0.01$) negative Pearson correlation coefficient of -0.34. For men, the correlation between jitter and $F_0$ was -0.43. For women, however, the correlation was insignificant, suggesting that absolute jitter did not appear to change as a function of $F_0$. The mean $F_0$ for each of the six semitone levels also was correlated with its corresponding mean absolute jitter, and they found strong negative correlations in men (-0.73, $p< 0.01$) and women (-0.88, $p< 0.01$). Orlikoff and Baken concluded that jitter levels are inversely related to the relative $F_0$ level within a speaker's phonational frequency range rather than to absolute $F_0$. Further they suggest that as $F_0$ increases, jitter in women's voices changes much less than it does in men's voices. It may be more meaningful to attempt to relate speaker's relative vocal $F_0$ values to jitter than to absolute $F_0$.

### Studies differentiating normals from pathological groups:

Many studies have "appeared to relate specific or general laryngeal diseases and the perception of vocal disorders to a number of perturbation measures"(Titze et al. 1987, p. 252). Hecker et al. (1971) analyzed recordings of the Rainbow Passage (Fairbanks, 1960) read aloud by 14 patients with cancer of the larynx. When each subject's data were matched with those of a normal speaker who was about the same age and exhibited a comparable average $F_0$, the pathological speaker's data exhibited a significantly higher directional perturbation factor. Murry and Doherty (1980) also were interested in differentiating normal speakers from those with cancer of the larynx. Both directional and magnitudinal perturbation factors discriminated 5 normal from 5 laryngeal cancer patients. Zyski, Bull, McDonald and Johns (1984) measured jitter in a group of 20 subjects with normal voices and a group of 52 subjects with a variety of laryngeal pathologies [(e.g., among those included were carcinoma, unilateral paralysis, spastic dysphonia, benign tumors (hyperkeratosis, nodules, polyps, papilloma)] and hoarseness without pathology. Results indicated that the best predictor of group membership was average pitch perturbation, followed by relative average perturbation, average amplitude perturbation and amplitude perturbation quotient.

Iwata and von Leden (1970) measured the range of pitch perturbations from the recorded vowel /a/ sustained for several seconds. Subjects were 30 normal speakers and 52 speakers afflicted with a variety of laryngeal disorders (20 with chronic laryngitis, 10 with benign tumors (6 vocal nodules, 2 hematomas, 1 leukoplakia and 1 papilloma), 6 with malignant tumors, 10 with unilateral vocal fold paralysis, and 3 with myasthenia laryngitis). Results revealed that 1) the pitch perturbations in the normal speakers ranged from -0.7 to + 0.6 ms, with slightly higher jitter values for the females than males, and exceeded the 0.5 ms jitter magnitude reported by Lieberman (1961); and 2) jitter magnitudes differed significantly between patients with laryngeal diseases and normal speakers, and among patients with different laryngeal pathologies. From these results Iwata and von Leden (1970) suggest that documenting changes in pitch perturbations may assist in providing an objective assessment of laryngeal dysfunction.

Ramig et al. (1988) applied acoustic analysis to the study of patients with various neurological diseases. The purpose of their study was three-fold. They sought to investigate the role of acoustic analysis in: 1) the diagnosis of diseases of different neural subsystems, 2) the measurement of disease progression, and 3) the diagnosis of subclinical manifestations of neurological disease. Patients with myotonic dystrophy (n=6, 5 men, 1 woman), Huntington's disease (n=8, 4 men, 4 women), and Parkinson's disease (n=8, all men) were selected to address the first question. One 69 year old man with amyotrophic lateral sclerosis (ALS) was studied over a six month period to address the second question, and 20 symptom-free subjects (6 men, 14 women) at risk for Huntington's disease were selected to study the third question. Seventy-eight subjects served as a control group, divided into three sub-groups to represent the age and sex distributions of the experimental groups. Audio recordings of subjects' productions of the vowels /a/, /i/ and /u/, sustained for "as long as possible" and at a comfortable frequency and intensity were digitized at 20kHz. The voice analysis program GLIMPES was used to measure $F_O$, percent shimmer and jitter and harmonics-to-noise ratio.

Results indicated that patients with distinct neural subsystem diseases exhibited distinct characteristics of voice. Ramig et al. (1988) compared the oscillograms of vowel phonation data from one patient representing each disease group with those of a control subject. This visual comparison revealed different oscillographic vocal patterns for each subject. The myotonic voice was characterized by relatively large percent shimmer, in contrast to the Huntington's disease patient who exhibited an abrupt phonatory disruption (revealed in an abrupt drop in $F_0$), rather than cycle-to-cycle deviances. The Parkinson's disease patient revealed rhythmic amplitude tremor imposed upon cycle-to-cycle deviances represented by increased jitter and shimmer magnitudes. Reported jitter and shimmer magnitudes showed that increased jitter values were most apparent for myotonic patients, with values 3 times greater than those reported for the normal controls. Ramig et al. (1988, p. 168 ) suggest that the elevated jitter values in the voices of the myotonic and Parkinson's disease patients are probably due to "the large cycle-to-cycle instability of the integrity of the muscles governing the vocal fold elasticity (VFE)." The acoustic findings also confirmed earlier perceptual work that classified the various voice disorders of the patients (Aronson, 1980).

A sample of sustained vowel phonation at the beginning and end of a six month period in the patient with ALS revealed an increase in phonatory instability characterized by an increase in percent shimmer. While percent jitter did not change appreciably over the six month period, shimmer and harmonics-to-noise ratio were most sensitive to deterioration, with values increasing from 2.18% to 3.33%, and decreasing from 21.74 dB to 16.59 dB, respectively. Data from this ALS patient support the hypothesis that "acoustic analysis reflected deterioration that accompanied neuronal degeneration in this patient, and support the sensitivity of acoustic analysis in reflecting early bulbar symptoms before they were clinically observed" (Ramig et al., 1988, p. 170).

Ramig et al. (1988) also observed from oscillograms an incidence of low frequency segments in sustained vowels approximately 12 times greater in patients at risk for Huntington's disease than in subjects of the control group. These findings support "the hypothesis of increased phonatory instability in certain individuals at risk for Huntington's disease" (Ramig et al., 1988, p. 170). No comment was made regarding jitter and shimmer magnitudes, or harmonics-to-noise ratio. Further studies are needed to evaluate the representativeness of these findings for each of the clinical groups.

Hirano, Hibi, Yoshida, Hirade, Kasuya and Kikuchi (1988) applied acoustic analysis to the study of patients belonging to three different disease groups. The purpose of their study was four-fold. They sought to determine the role of acoustic analysis in: 1) differential diagnosis of different disease groups, 2) monitoring the effects of treatments, 3) relation to other clinical voice tests and 4) determining if there are correlations among the acoustic parameters. Acoustic analysis was performed on the tape-recorded voice samples of 30 patients before and after treatment: Ten had glottic TIa carcinoma (mean age = 64 years), ten unilateral vocal fold polyps (mean age = 46 years) and ten unilateral recurrent laryngeal nerve paralysis (mean age = 59 years). Therapy consisted of laser surgery with/without radiotherapy for the carcinoma cases, endolaryngeal microsurgery for the polyp cases and intrafold silicone injection for the paralysis cases. Three acoustic measurement parameters were employed: Pitch Perturbation Quotient (PPQ), Amplitude Perturbation Quotient (APQ) and normalized Noise Energy (NNE). Results indicated that PPQ and APQ were greater in paralysis cases than in carcinoma and polyp cases. However, these acoustic measurement parameters were not useful in differentiating the disease groups studied. In addition, the range of values for each of the groups overlapped and a combination of the acoustic variables did not differentiate the three groups. Hirano et al. speculate that differential diagnosis seems to depend on what kind and degree of pathology one is working with. The three acoustic parameters did provide useful information for monitoring the effects of treatment; the majority of patients in whom voice was improved following surgery also showed a decrease in PPQ, APQ and NNE values. In those cases in which the voice did not improve following surgery, the PPQ, APQ and NNE values increased. The acoustical measures (PPQ, APQ and NNE) were highly positively correlated among each other and with judgements of hoarseness, rough vocal quality and breathy voice quality as measured by a psychoacoustic scale. These authors did not specify the nature of the speech sample used (vowel vs. connected speech), the sampling rate or hardware used for analysis.

Ramig, Scherer, Klasner, Titze and Horii (1990) examined acoustic measures of phonatory instability (coefficient of variation for amplitude, coefficient of variation for fundamental frequency, jitter, shimmer and harmonics-to-noise ratio) in the sustained vowels /a/ and /i/ produced by a 69-year-old man diagnosed with amyotrophic lateral sclerosis (ALS), a sex- and age-matched control subject, and on one occasion, a 16-subject control group. These measures were taken five times throughout a six month period. Recorded voice samples were digitized at 20 kHz via a 16-bit A/D converter and analyzed using custom software (Scherer et al., 1988). The last recording of the ALS patient was characterized by an increased coefficient of variation for amplitude, an increased coefficient of variation for frequency, increased jitter and shimmer, reduced harmonics-to-noise ratio as well as reduced maximum vowel duration. Although the ALS patient's jitter increased across trials for the six month period, these jitter data still fall within the range of jitter for the control group and the longitudinal control subject. This acoustic evidence of phonatory instability was in contrast to the consistency of the acoustic measures in the control subject and control group. Ramig et al. (1990) interpret their findings as support for the usefulness of monitoring disease progression by means of acoustic analysis.

Pruszewicz, Obrebowski, Swidzinski, Demenko, Wika and Wojciechowska (1991) measured formant frequencies, fundamental frequency and its range, percentage of noise as well as mean and maximum values of jitter in sustained vowels (/e/, /i/) produced by thirty patients with dysphonia. These included fifteen patients with organic type and fifteen patients with functional type dysphonia. Jitter measures were derived from spectrograms. The authors did not specify the length of the vowel sampled, the perturbation formulae or sampling rate used in their study. Results indicated that jitter was significantly greater (8%) in the organic group than jitter measured in the vowels of the functional group (4%). The speech signal demonstrated a less disfigured harmonic structure in functional voice disorder patients than in patients with organic type dysphonia. These authors found broadband noise disfiguring the signal; actual values obtained for fundamental frequency and its range as well as percentage of noise in the signal were not reported. Pruszewicz et al. suggest that these data be accepted as pilot studies and that future investigations should explore the possibility of using selected acoustic methods to aid in the differential screening of functional versus organic type dysphonia due to laryngeal pathology.

The studies reviewed emphasize the need for further investigations exploring the value of acoustic analysis for early detection and differential diagnosis of laryngeal disorders, as well as for monitoring disease progression and response to rehabilitation. These measures will complement the perceptual skills of the speech-language pathologist by providing more sensitive and objective data on the functioning of the laryngeal mechanism.

## Perceptual Studies

A limited number of studies have examined the relationship between objective measures of jitter and perceptual judgements of voice quality. Isshiki, Yanagihara and Morimoto, (1966) and Yanagihara (1967) suggested that two components of "hoarseness" are evident in a "real" voiced speech signal and these may influence perceptual judgements of voice. One component is the noise associated with period and amplitude fluctuations of the glottal wave and the other is due to the leakage of air through the glottis without noticeable perturbations of the glottal waveform. Early investigators were interested in the relationship between artificially-jittered synthetic stimuli and perceived hoarseness or roughness in such signals (Coleman, 1969; Wendall, 1963).

Coleman (1969) investigated the relationship between median frequency levels and roughness perception for jittered stimuli. Twenty stimuli were produced by a computer program (LADIC) at five different frequency levels and four jitter excursions. Results showed that stimuli of lower median frequency were judged by 32 psychology students (16 men and 16 women) to be rougher than those of higher median frequency for all jitter conditions.

Wendall (1963) was interested in the degree of aperiodicity necessary in synthetic stimuli for listener judgements of roughness. The same computer program (LADIC) that was used by Coleman (1969) was used to generate stimuli that varied in the size of frequency differences between successive cycles. For each of two median frequency conditions, 100 and 200 Hz, stimuli were generated to have frequency variations of ± 10, 8, 6, 4, 2, and 1 Hz. Five hundred and thirty-five listeners served as judges; results indicated that frequency variations as small as ± 1 cycle around a median frequency of 100 Hz may be perceived as rough. A direct relationship between the degree of perceived roughness and the frequency differences between successive cycles was found for a given median frequency value. A frequency variation of ± 10 Hz was judged to be more rough than a ± 1 Hz variation around a median frequency of 100 or 200 Hz. When the stimuli were placed in rank order, the 200 ± 10 Hz signal was judged to be less rough than the 100 ± 2 Hz stimulus. This suggests that if a listener hears a male and a female voice with the same frequency variation, the male will be judged to have the more deviant voice, at least with respect to vocal "roughness."

Rozsypal and Millar (1979) investigated the perceived naturalness of synthetic vowels /a/, /i/ and /u/ with controlled amounts of introduced jitter and shimmer. These authors synthesized three-formant tokens of the vowels, introducing different quantities of jitter and shimmer at the source. An analog speech synthesizer (PAT) programmed five levels of jitter that ranged from 0 to 3.20 % and five levels of shimmer that ranged from 0 to 15.19 %. In a paired-comparison experiment twenty-seven subjects were asked to rate the naturalness of the vowels presented once over a loudspeaker using a seven point equal-appearing interval scale. Results indicated the optimal amounts of jitter to be 0.80% for /u/, 1.60% for /a/ and the optimal amount for /i/ ranging from 0 and 1.60%. The presence of shimmer always decreased naturalness. No trade-off relationship was found between jitter and shimmer for perceived naturalness. That is, they did not find that jitter and shimmer had equivalent and additive effects on the perception of naturalness. However, interpretation of these results is made difficult because jitter and shimmer were not independent in the stimuli. This study also did not consider the effect of acoustic interactions such as jitter, shimmer and SNR in the vocal tract.

Hillenbrand (1988) explored the relationship among jitter, shimmer and additive noise in the perception of synthetically generated voices. Previous studies have reported significant intercorrelations among individual acoustic parameters such as jitter, shimmer and additive noise (Davis, 1976; Deal & Emanuel, 1978; Heidberger & Horii, 1982; Horii, 1980). Hillenbrand (1988) found that strong relationships existed between perceived roughness and variation in either pitch or amplitude: the perception associated with jitter was different from that associated with shimmer, and a strong relationship was found between the perception of breathiness and additive noise. In addition, Hillenbrand (1987) found that presently-available acoustic analysis techniques are not always able to discriminate among the different sources of aperiodicity. This suggests that studies investigating the relationship between jitter and the perception of hoarseness also should include shimmer and signal-to-noise ratio among their acoustic measures. The question remains as to how these acoustic parameters combine to influence listener judgements of the overall severity of a dysphonia.

Other investigations have explored the relationship between natural phonation and perceived roughness (Kane & Wellen, 1985; Deal & Emanuel, 1978; Askenfelt & Hammarberg, 1986). Kane and Wellen (1985) collected pitch and amplitude perturbation measures from 10 children with vocal nodules during sustained /a/ at a comfortable pitch and loudness level and compared these measures to clinical ratings of severity of vocal quality disorder using the Wilson Voice Profile system (1971). Thirty seconds from the middle of a passage read aloud or spontaneous speech sample produced by the children were chosen for rating. Perturbation values were obtained from the speech signal rather than the residue signal. Results indicated that pitch perturbation quotient (PPQ) measures ranged from 0.0023 to 0.0472 ms with a mean of 0.012 ms. Jitter correlated significantly with the Wilson ratings ($r=.68$, $p< .05$). Unfortunately, the authors measured jitter in sustained vowels, and the perceptual ratings were made on the basis of connected speech. Therefore the perceptual ratings cannot be assumed to be comparable to the jitter measurements.

Deal and Emanuel (1978) measured jitter and shimmer associated with vowel phonations representing a range of vocal roughness. Subjects were 20 men with normal voices and 20 men with hoarse voices and diagnosed laryngeal pathology. The normal subjects were asked to produce 5 vowels /u, i, ^, a, ae/ with their habitual voices and then with voices that simulated roughness. Deal and Emanuel found that increases in vowel acoustic wave variability (estimated by jitter and shimmer) were associated with increases in vowel spectral noise levels and perceived vowel roughness for the normal, simulated rough and clinically hoarse productions. However, the size of the relationship observed between shimmer or jitter and perceived roughness did not seem large enough for either to be used as a reliable, clinical predictor of roughness. This may partially be explained by the fact that noise associated with period and amplitude fluctuations of the glottal wave as well as the noise due to air leakage through the glottis may have influenced perceptual roughness ratings. A positive, nearly linear, relationship has been observed between spectral noise levels and perceived roughness. They suggest that spectral noise may provide a more clinically useful roughness indicator than the jitter and shimmer indices used.

Askenfelt and Hammarberg (1986) related their objective acoustic measures of waveform perturbations in a reading passage produced by 41 patients with voice disorders, to perceptual ratings of voice made by six speech pathologists for the same patients before and after voice therapy. The majority of subjects (33) presented with organic disorders, predominantly mass lesions of the vocal folds (i.e., polyps, nodules, carcinoma, and papilloma). Askenfelt and Hammarberg compared the results of seven different waveform perturbation formulae applied to tape-recorded running speech samples from the 41 voice patients, before and after successful voice therapy. The acoustical analysis did not distinguish between pitch and amplitude variations. Four speech pathologists were first asked to provide an " overall rating of the degree of perceptual deviance from normal voice using a 5-point equal appearing interval scale" (Askenfelt & Hammarberg, 1986, p. 52). A second perceptual rating performed by three of the same speech pathologists focussed on qualities associated with roughness/instability. Judges were asked to rate voices in their own terms. Ten of the twelve most frequently used terms were then selected and used to rate the degree of deviance from normal voice on the same 5-point scale as in the preceding "overall" rating. Correlations between each of the seven acoustical signal perturbation measures and the perceptual judgement ratings were compared with regard to their ability to distinguish between normal and pathological voice. The correlation coefficients between the perceptual ratings and the acoustic measures were plotted. Askenfelt and Hammarberg found that the $DF_0$ ( the standard deviation of the distribution of the relative frequency differences) provided the best correlation. Unfortunately, the authors measured jitter in the voiced segments of running speech, and the perceptual measures were based on the entire running speech samples. Therefore similar to the design limitations on the data of the Kane and Wellen (1985) study, the perceptual ratings obtained by Askenfelt and Hammarberg cannot be assumed to correlate directly with the jitter measurements. The majority of the perceptual voice parameters showed significant improvement between pre- and post-treatment recordings. Raw vocal jitter and shimmer values for each of the patients before and after therapy were not presented, however.

Zajac and Linville (1989) investigated the relationship between jitter and shimmer and perceived nasality in voices of children diagnosed with velopharyngeal incompetency (VPI) without any vocal pathology. Subjects were ten children with VPI and five normal children who served as controls. Subjects were asked to produce the vowels /i/, /a/, and /u/ for approximately three seconds at a comfortable loudness level. They were also asked to produce sentences containing nasal consonants. Jitter, shimmer and fundamental frequency were measured from the midportion of each vowel. A computer program using a zero-crossing algorithm detected jitter and shimmer from an EGG signal. The perceptual ratings were based on the sentences using a seven-point equal interval scale for nasality and a five-point equal interval scale for rating hoarseness. Results indicated that jitter was significantly higher in children with VPI than those in the control group. Differences in shimmer were non-significant between the two groups. A positive correlation was found between shimmer and perceived hoarseness. In contrast jitter was found to be positively correlated with listeners ratings of perceived nasality. Unfortunately similar to the methods of Kane and Wellen (1985) and Askenfelt and Hammarberg (1986), Zajac and Linville (1989) also based their acoustical and perceptual analyses on two different contexts.

Wolfe, Cornell and Palmer (1991) sought to explore the relationship between acoustic parameters (harmonic/noise differences representing selected portions of the spectrum, average fundamental frequency, natural logarithm of fundamental frequency and jitter) and different voice types (hoarse, breathy, strained, and normal). Subjects included 20 males and 31 females referred for voice therapy. Subjects were asked to produce the vowels /a/and /i/ for several seconds. Measures of mean harmonic noise difference were obtained from spectograms; average fundamental frequency and jitter were obtained by presenting tape recordings to a Kay Electric Visi-Pitch (Model 6095) for acoustical analysis. The length of the recorded vowel sample for acoustic analysis was 100 ms. The sampling rate for jitter analysis was not specified. Listeners were 11 undergraduate speech-language pathology students. One-second vowel samples were spliced together to form a master listening tape for each of the subjects. Following two 30-minute training sessions, listeners were asked to classify each vowel as "primarily normal," "primarily breathy," "primarily hoarse," or "primarily strained." Analysis of the data included: 1) providing a description of the different voice types based on spectral curves, 2) using direct discriminant analysis to determine the ability of acoustic variables to discriminate among different voice types, 3) using a stepwise discriminant analysis to determine the best combination of variables for predicting voice type and finally, 4) determining the number and type of misclassifications based upon discriminant analysis. Results of their study indicated that classifications of voice type were made with 80 % accuracy (37 out of 46 predictions) using three acoustic parameters. These included mean harmonic/noise difference factor (MNDF), natural log of fundamental frequency ($LNf_0$) and vowel type. These authors suggest that spectral information representing second and third vowel formants in combination with fundamental frequency may be especially useful in predicting voice types. The specific contribution of a particular vowel could not be determined from their analysis. Jitter magnitudes did not contribute to the classification of voice type. However, Wolfe et al. (1991) suggest that other measures of period variation might prove more useful in the prediction of perceived vocal roughness.

Feijoo and Hernandez (1990) investigated the vocal quality of 64 normal subjects and 57 subjects afflicted with various degrees of glottal cancer using six different acoustic measures of a voiced signal. The acoustic measures included tone period perturbation, amplitude perturbation, waveform perturbation, vocal noise, spectral periodicity (Cepstrum) and spectral distortion. Subjects were asked to sustain the vowel /e/ five times at a comfortable pitch and loudness level. The recorded vowel signals were first low-pass filtered at 4.8 kHz and then sampled at 40 kHz. Autocorrelation was employed to determine the tone period of the voices. The acoustic measures were compared to perceptual evaluations of " grade" and "breathiness" performed by 4 trained listeners. "Grade" was defined as the overall degree of hoarseness in the signal, and "breathiness" referred to the vocal quality produced by turbulent noise resulting from incomplete glottal closure. Listeners rated each of the subject's five vowel utterances according to breathiness and grade separately using a four-point scale with 0 representing a nonpathological voice at one end and 3 a severely pathological voice at the other end of the scale. The six acoustic measures were calculated for the normal control group and 50 of the glottal cancer group. Results of classifying the subjects based on their acoustic parameters revealed that cepstrum, vocal noise, and spectral distortion had the smallest probability of error. Amplitude perturbation, waveform variability and pitch perturbation had the highest probabilities of error with 14.03, 14.03, and 22.81 %, respectively. The results of classifying subjects based on the perceptual ratings revealed similar misclassification rates as those based on the acoustic parameters (12.28 %). Correlations between the acoustical parameters and perceptual judgements were best for amplitude perturbation (0.915 with breathiness and 0.930 with grade) and the worst for tone period perturbation ( 0.750 with breathiness and 0.754 with grade). Similar to the suggestion by Hillenbrand (1988), Feijoo and Hernandez (1990) suggest that a combination of acoustic parameters may discriminate more effectively between healthy and sick subjects than reliance on a single acoustical measure alone.

Eskenazi, Childers and Hicks (1990) also explored the relationship between acoustic measures in the vowel /e/ produced by 25 males and 25 females with no vocal disorders and 16 patients with vocal disorders , and perceptual ratings of the same 2-second voice signal made by seven listeners . Six acoustic measures were examined. These included spectral flatness of the residue signal (SFR), pitch amplitude (PA), harmonics-to-noise ratio (HNR), Pitch Perturbation Quotient (PPQ), Amplitude Perturbation Quotient (APQ) and percent jitter. Perturbation data were obtained from EGG signals that were low-pass filtered at 4800 Hz and sampled at 10 kHz. Vowel sampling included voice onset and offset characteristics. Listeners were asked to rate overall severity, hoarseness, breathiness, roughness, vocal fry and overall excellence of the 50 normal voices on a 7-point scale. Multiple linear regression analysis between the acoustic measures and the perceptual judgements revealed that the two most useful acoustic parameters for predicting vocal quality in the voice-disordered subjects were PA and the HNR. Two of the measures, PPQ and APQ were not included in the listeners evaluation of pathological speakers voices. No single acoustic measure was useful in predicting vocal quality in the normal subjects. Judges relied on different parameters and Eskenazi et al. further suggest that judges may be relying on other parameters and that the degree of reliability for predicting vocal quality may be improved by adding other acoustic variables not used in their study.

The studies reviewed suggest that there is a relationship between acoustic measures and voice quality. Unfortunately the acoustic measures and perceptual judgements have not always been based on the same context. Further studies using the same context (sustained vowels or connected speech) for acoustical and perceptual analysis are needed to clarify the relationship between different acoustic measures and voice quality.

In sum, one must consider and control for measurement, task and speaker variables in interpreting the results of vocal jitter studies. A number of studies indicate that significant correlations exist between jitter values for synthetic- and naturally-produced voiced stimuli and perceptual judgements of the quality of those stimuli. However some of these studies exploring the relation between natural phonation and perceived roughness have failed to base their perceptual ratings on the same context (i.e. material spoken, vocal mode) (Kane & Wellen, 1985; Askenfelt & Hammarberg, 1986; Zajac & Linville, 1989). Others have not specified the basis on which their perceptual ratings were made (Deal & Emanuel, 1978). To date only one study has looked specifically at objective acoustic measures of perturbation and correlated these with perceptual judgement ratings before and after voice therapy (Askenfelt & Hammarberg, 1986). This study did not separate the two types of perturbation, jitter and shimmer, or present the perturbation values for each of the subjects. Studies attempting to accurately measure jitter and correlate these values with perceptual measures also should include shimmer, signal-to-noise ratio (SNR) and fundamental frequency among their acoustic variables. Studies have shown that these measures covary, and it may not be possible to measure the effects of one independently of the effects of the others on the perception of voice (Hillenbrand, 1987; Hillenbrand, 1988). As well studies should consider the influence of noise associated with period and amplitude fluctuations of the glottal wave versus noise due to air leakage through the glottis on the perceptual evaluation of voice.

With these points in mind, the principal investigator undertook the present study to explore the relationship between acoustical and perceptual voice analysis methods. This research has attempted to address some of the acoustical analysis methodological issues described in the preceding text by controlling for some of the variables that are thought to affect jitter but have been uncontrolled in previous studies of this kind. Specifically, jitter and shimmer magnitudes, fundamental frequency ($F_0$) and signal-to-noise ratio (SNR) were measured in the sustained vowel /a/ produced by four adult females with vocal nodules, before and during voice therapy. The relationship between these acoustical measures and listeners' perceptions of dysphonia in the subjects' voices before and during the last stage of treatment were also measured. Specifically the following questions were addressed:

1) Is there a difference in jitter magnitude levels for sustained productions of /a/ before, and during voice therapy in four adult females with vocal nodules?

2) What is the relation among jitter, shimmer, $F_0$ and SNR magnitudes in paired samples of sustained /a/ obtained during the last week of baseline and the last week of treatment ?

3) Is there a perceptual difference in severity of dysphonia between paired samples of sustained /a/ obtained during the final week of baseline and the final week of treatment?

4) Is the jitter magnitude difference between paired samples of sustained /a/ obtained during the final week of baseline and the final week of treatment correlated with the perceived difference in severity of dysphonia between the same pairs of /a/ samples?

5) What are the relative contributions of each of the acoustic variables (jitter, shimmer, $F_0$ and SNR) to the perception of change in voice quality?

# CHAPTER III. METHODOLOGY

## Design

This research took the form of a single-subject A-B and correlation design, implemented on four subjects. The A-B components of the study allowed for pretreatment tracking of the dependent variables and documentation of their change across a period of treatment (McReynolds & Kearns, 1983).

Experimental Variables. The independent variable chosen to effect a change in voice was voice therapy, and there were five dependent variables: four acoustical ( jitter, shimmer, signal-to-noise ratio (SNR) and fundamental frequency ($F_0$)), and one perceptual measure (ratings of dysphonia).

Experimental Phases. The A-B components of the proposed research consisted of two experimental phases: baseline and treatment. The baseline (A) was four weeks in length, and the treatment phase (B) was eight weeks in length with the exception of the treatment phase for subject two who requested a shortened treatment period; her baseline phase was four weeks in duration, and treatment lasted six weeks. During each phase data were collected three times per week for the four acoustic dependent variables.

The presence of vocal nodules was documented in each of the subjects via a flexible fiberoptic video nasendoscopic (FFVN) laryngeal examination performed by a speech-language pathologist at the University of Alberta. A diagnosis of vocal nodules for each of the subjects was substantiated by an otolaryngologist who reviewed the pre-treatment FFVN records.

1) Baseline

During the four-week baseline condition, each subject's voice was recorded as she produced /a/ sustained at comfortable pitch and loudness levels for approximately five seconds. Baseline recordings were obtained three times per week across the four weeks. No treatment was provided during this time.

2) Treatment

The treatment phase consisted of a six-to-eight week period during which voice therapy was provided to the subject three times per week by the principal investigator in the role of the clinician. Probe voice recordings for the acoustical dependent variables continued to be obtained at the outset of each treatment session for the duration of this phase.

3) Follow-up

At the end of the treatment phase, a second flexible fiberoptic video nasendoscopic laryngeal examination performed by a speech-language pathologist at the University of Alberta provided an opportunity to document any structural change in the subjects' vocal nodules. An otolaryngologist provided a second opinion to substantiate any such change in size or disappearance of the vocal nodules following treatment upon viewing the end-of-treatment FFVN videotapes. During the few days in which subjects waited for the post-treatment FFVN evaluation, they continued to come in for probe recordings of the acoustical dependent variables, although no treatment was provided and no verbal feedback regarding voice techniques, improvement or relapse was given during these post-treatment visits.

The experimental design was organized into phases divisible by four weeks (with the exception of the treatment period for Subject two) for several reasons. It was expected that a four-week pretreatment phase would provide an opportunity to establish a stable baseline and observe the effects of factors which may influence jitter magnitude such as the menstrual cycle and practise effects. It was also expected that the six-to-eight week treatment phase would allow sufficient time for demonstration of a change in voice with voice therapy and provide further opportunity to monitor menstrual cycle effects, if any, on the dependent variables.

## Subjects

Potential subjects were adult females diagnosed with vocal nodules at the Voice Clinic at the Glenrose Rehabilitation Hospital and at the University of Alberta. The research study was suggested to these patients as an alternative to waiting for treatment at the Glenrose, the University or local Health Unit. Interested participants were asked by clinic personnel to sign a form allowing the primary investigator to contact them by telephone and provide further information regarding the study. (See documents in Appendix A.)

Eligible subjects must have abstained from smoking for at least two years, have no history of respiratory illness, chronic allergies, or speech and hearing problems. Subjects must also have abstained from taking drugs (e.g., antipsychotics, cortisone, antidepressants, antihistamines) for two weeks prior to participating in the study. All of these are factors which have known or suspected effects on the voice (Gilbert & Weismer, 1974) and may negatively influence the success of voice therapy.

Finally, all subjects were required to pass a pure-tone hearing screening at 35 dBHL for frequencies 500, 1000, and 2000 Hz (ASHA, 1978). This was a necessary screening procedure, because severe hearing loss over a long period of time can result in a change in habitual vocal pitch levels and pitch variability (McGlone & Hollien, 1963). Hearing loss may also influence success in voice therapy (Boone, 1983).

Participants who met all of the inclusion and exclusion criteria were four females between 25 and 40 years of age. All subjects were fluent speakers of English; two of the subjects were native speakers of western Canadian English; the other two reported Japanese and Chinese to be their first languages. A summary of subject information is presented in Table 1.

Table 1 Subject Information

| | **Subject One** | **Subject Two** | **Subject Three** | **Subject Four** |
|---|---|---|---|---|
| **Speaker** | Japanese Canadian | English Canadian | Chinese Canadian | English Canadian |
| **Age** | 31 | 25 | 40 | 30 |
| **Occupation** | homemaker/ mother of 3 young children | homemaker/ trained singer/ mother of triplets | restaurant & bar owner/ mother of 2 young children | fitness instructor consultant |
| **General health** | good | monthly colds with onset of menses | poor, recurrent colds | good |
| **Reported duration of voice problem** | > one year | ~ nine months | ~ six months | ~ six months |
| **Pathology** | 2-3 mm bilateral vocal nodules (larger on the right fold) | 2-3 mm bilateral vocal nodules | 3-4 mm bilateral vocal nodules | pinpoint bilateral vocal nodules with evidence of scarred tissue on the right vocal fold |

Materials / Equipment

Materials and equipment required for initial subject selection included case history questionnaires, and a portable audiometer (Maico MA-41, calibrated to ANSI-1969 standards). Acoustical data acquisition and analysis tools included reel-to-reel direct audio tape recordings transduced by a uni-directional dynamic microphone, and filtered for digital waveform analysis via the CSpeech program (Milenkovic, 1989, cited in Read, Buder & Kent, 1990) on a laboratory-dedicated microcomputer (IBM-PC-AT clone). Perceptual data acquisition and analysis materials included reel-to-reel direct audio tape recordings filtered as for digital analysis and played back to listeners via a loudspeaker system.

## Procedures

Subject selection. An initial screening of potential research subjects was conducted by the primary investigator over the telephone (Appendix B). Questions were asked regarding the candidate's history of speech and hearing problems, smoking, respiratory illness, neurological disease and allergies. This procedure attempted to eliminate unnecessary trips to the research site at the University of Alberta for further screening procedures by candidates who were ineligible on the basis of the preceding criteria. Following the telephone screening, the vocal nodules treatment information form was mailed to potential subjects for their perusal (Appendix C).

Subject participation extended over 10 -12 weeks with three sessions per week. At the beginning of the first session subjects were re-apprised of the purpose and procedures of the study, and their written consent was obtained (Appendix D). A voice case history was obtained to ensure that subjects did meet all selection criteria (Appendix E), and subjects' hearing was screened. Because the menstrual cycle is thought to have effects on the pitch and pitch variability of the voice (Higgins & Saxman, 1989), baseline and treatment phases each incorporated at least one complete menstrual cycle in an attempt to control for its possible effects on fundamental frequency and jitter values before, and during therapy. Subjects were asked to notify the primary investigator when they began their menstrual cycles. The influence of colds/allergies was also tracked over the baseline and treatment sessions. Subjects were required to participate consistently at a pre-arranged time between noon and 3 p.m., or in the evening, to control for any variation in the voice that might be associated with time of day (Boone, 1971). This criterion was applied throughout the study.

Voice data recordings. The instrumental array for voice data acquisition is illustrated in Figure 1. Baseline recordings of subjects' sustained productions of the vowel /a/ were obtained during the first and subsequent sessions over the first four weeks. Each of these recording sessions took approximately 15-20 minutes. Probe recordings of the same vowel continued to be obtained across the six-to-eight weeks of the treatment phase. This resulted in 108 probe recordings (3 repetitions of the vowel /a/ x 3 times/week x 12 weeks) for subjects one, three and four, and 90 probe recordings (3 repetitions of the vowel /a/ x 3 times/week x 10 weeks) for subject two.

For each of these probe recordings subjects were seated in a noise-controlled environment and fitted with a head-mounted, uni-directional, dynamic microphone (Shure, SM 10A) aligned 13 mm from the mouth. To ensure that the voice signal was recorded at a consistent level, the VU meter on the direct reel-to-reel tape recorder (Sony, TC-788-4) was monitored by the investigator. To control for mechanical influences of the tape drive associated with tape load on jitter magnitude levels, the investigator prepared equal and individual lengths of tape for each and every probe recording obtained from each subject. A subject was required to sustain phonation of the vowel /a/ three times for five seconds each time at her habitual pitch and a comfortable loudness level. A demonstration by the clinician and three practice trials were given prior to recording the samples used for analysis to ensure a subject performed to the best of her ability. To encourage each subject to use her habitual pitch level, the primary investigator instructed her to say the word "hello" concentrating on the pitch of the voice during the "o" and using this pitch when she produced each sustained vowel. The fundamental frequency of the "o" at the end of the word is thought to approximate habitual pitch.

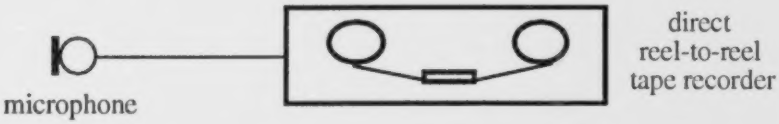


Figure 1. Instrumental array for voice data acquisition.

Although the use of tape recorders for storing data for later acoustic analysis has been known to slightly elevate jitter and shimmer levels (Doherty & Shipp, 1988), a recorded signal was chosen for the following reasons: 1) tape recordings are most often readily available to and used by clinicians to monitor changes in voice disorders associated with therapy; 2) Doherty and Shipp (1988) found that although direct sampling or the use of a video cassette recorder with pulse code modulation provide the most sensitive measures of jitter and shimmer magnitudes, "any good quality tape recorder may be adequate for studying voices that would be expected to exhibit jitter and shimmer magnitudes above normal levels" (p. 489) ( e.g., as in the case of voices affected by vocal nodules); and 3) it was important that the same voice signal be delivered to the computer for acoustic analysis and to the listeners for perceptual ratings. If a live signal had been used for digital analysis, perceptual ratings would have to have been made from playback of a reconstituted (digital-to-analog) recording. This could have introduced the possibility of further distortion in the dependent variables of interest or the introduction of artifacts in unknown amounts in the reconstituted (D/A) signal. The decision made by the primary investigator to use a recorded signal may have distorted the analog signal slightly but it would have been distorted in the same manner for computerized and human analysis.

Voice treatment. During the treatment phase of the project, a series of voice therapy tasks was presented to the subject by the clinician, and contingent reinforcement of the subject's performance was provided, using a 100% fixed reinforcement schedule. Following a subject's accurate productions of a vocal technique, the clinician provided verbal praise (e.g., "right," "good," "that's what I want to hear," etc.). When the subject failed to use a specific technique accurately in a word or phrase, the clinician provided a verbal reproof (e.g.,"no, that wasn't quite right"). This was followed by a demonstration of the correct production by the clinician and an explanation to the subject about what she was doing wrong and what she could try in order to obtain the target response. The treatment protocol over the six-to-eight weeks consisted of six components:

1) Explanation of the nature of the vocal nodules problem from physical, physiological and acoustical standpoints. This was explained by the primary investigator with the aid of pictures, diagrams and typed materials.

2) Explanation of vocal hygiene. Subjects were provided with an explanation of, and typed information about, good vocal hygiene habits.

3) Explanation, identification and elimination of vocally abusive behaviors. These were recorded daily by the patient on a chart and monitored over time. The chart was reviewed at the beginning of each session.

4) Demonstration and practise of "chewing" exercises to train the subject to use exaggerated movements of jaw, lips and tongue. This is thought to indirectly reduce tension in the voice. These exercises, as outlined in Boone (1971), were used at the beginning of each treatment session to relax the subject's articulators and aid in the production of a more relaxed voice.

5) Imagery to train the subject about "placement of the voice in the oral-facial mask." This technique, as outlined by Boone (1971), was designed to reduce strain in the laryngeal area, help improve vocal quality and thus produce a more efficient voice. It also is thought to indirectly reduce "hard glottal attack" and change pitch to a more natural level (D. Broda-Masiuk, personal communication, March 1990). The imagery of "placing" the voice in the midportion of the face was illustrated by means of diagrams and gestural cues provided by the clinician. "Placement of the voice" was first introduced through humming exercises during which the subject was encouraged to concentrate on the sound vibrations felt in the mid-facial area beneath the eyes. A treatment hierarchy for applying this imagery proceeded from humming to "placing the voice in the mask" in monosyllabic words containing nasal consonants, to multi-syllabic words containing nasal consonants, to short oral and nasal phrases, to oral and nasal paragraphs and finally to "placing the voice" during conversation. The patient was successful on 90% of a set of 20 imagery training trials before moving on to each subsequent level of the treatment hierarchy. Auditory judgements by the clinician and subject determined successful and unsuccessful trials. Consensus involved an agreement between the subject and clinician that there was an audible improvement in the voice as well as the subject's testimony that she felt less tension and effort while producing voice.

6) Elimination of "hard glottal attack". Subjects were encouraged to develop "easy phonation" with elimination of abrupt voice onset through exercises progressing from isolated vowels to single words, to phrases, and finally to conversation. A 90% accuracy rate over a set of 20 voice onset training trials was required before moving on to each subsequent level of the treatment hierarchy. As in technique #5, consensual auditory judgements of the subject's voice onset characteristics by the clinician and the subject were used to determine successful and unsuccessful trials.

A copy of the materials provided to subjects during treatment is in Appendix F.

Acoustic Data Acquisition and Analysis.

Calibration

The validity and reliability of the CSpeech program for fundamental frequency and perturbation analyses have been established for synthetically-jittered and normal voice signals (Milenkovic, 1987, 1989, cited in Read et al. 1990). Its validity was demonstrated for the proposed project in the following manner. A function generator (BK Precision 3011) was used as a source of 1000, 500 and 250 Hz sinusoidal, square and triangular waveforms. The integrity of its output was confirmed against a digital storage oscilloscope (Gould OS4200), low pass filtered at 10 kHz (Frequency Devices 901) and digitized at 20 kHz by the CSpeech software program via a 12-bit I/O board (Data Translation 2821) on an IBM-PC-AT clone (Zenith 286). Accuracy of the CSpeech program for digital frequency analysis of these waveforms was found to be within ± 1.0 Hz. The direct audio reel-to-reel recording system used in this research project (Sony TC-788-4) was calibrated for frequency and intensity according to the manufacturer's specifications. The frequency analysis reliability of the CSpeech program was tested by recording a sustained /a/ produced by the primary investigator via the head-mounted microphone on the Sony recorder and re-digitizing this signal three times. Reliability of the CSpeech program for this repeated frequency analysis was found to be 100%. Repeated jitter, shimmer and SNR analyses of the same recorded signal were also found to be 100% reliable.

Analog-to-digital conversion and analysis. The instrumental array for digital analysis of the recorded voice data is illustrated in Figure 2. (The D/A board, amplifier and loudspeaker portions of the array were used by the primary investigator to monitor the digitized signal for each of the vowels during the data analysis process.) The approximate mid three seconds of subjects' five-second phonation samples for the vowel /a/ were used for acoustical analysis. The three-second length of the analysis window meets the requirements suggested by Titze et al. (1987) and Karnell (1991). The first and last seconds of the signal were eliminated to avoid the influence of voice onset and offset characteristics on the acoustical analysis. The mid three seconds of the recorded vowel were low-pass filtered (8 kHz) to preclude aliasing and digitized at 20 kHz by the CSpeech program (version 3.0) on an IBM-PC-AT clone via a 12-bit analog-to-digital converter (DT 2821), yielding a temporal resolution of 50 µs (SNR about 96 dB). These samples were used for calculations of average $F_0$, jitter and shimmer magnitudes and SNR.

Jitter and shimmer analysis were performed using an interpolated short-term autocorrelation function described by Milenkovic, (1987; 1989; cited in Read et al. 1990). Jitter measures were based on the averages of complete pitch period cycles and expressed as the difference between estimated values of two successive periods (see Milenkovic, 1987, Eq. 16). Shimmer was expressed as the relative difference between amplitude estimations of two successive periods (see Milenkovic, 1987, Eq. 17). Finally, SNR was given as a 10 logarithm of the ratio between speech signal energy and the minimum energy of the error signal (see Milenkovic, 1987, Eq. 18). The autocorrelation was begun with an initial pitch period estimate performed manually by the investigator, and its variation was tracked automatically via a moving observation interval ($\leq$ one pitch period). The temporal resolution of the perturbation analysis was improved by parabolic interpolation. Concern has been raised regarding the tendency of the autocorrelation method to add or subtract formant oscillations in voice signals. These errors tend to occur at transitions between vowels or between vowel and consonant combinations, however, and Milenkovic states that "the autocorrelation method of pitch extraction is sufficiently reliable in measuring the pitch of sustained vowels" (1987, p. 531).

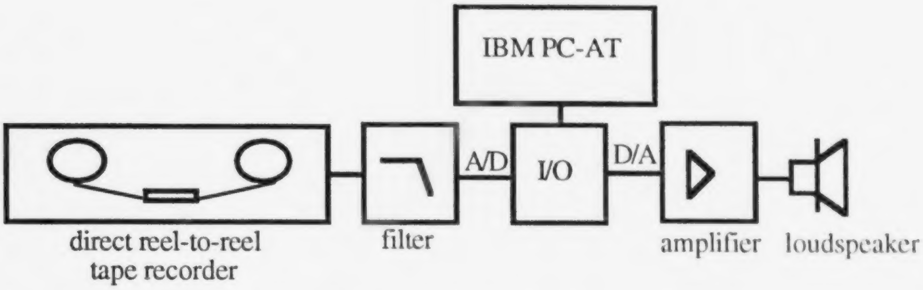


Figure 2. Instrumental array for digital analysis of the acoustical data.

Perceptual Data Acquisition. The instrumental array for perceptual data acquisition is illustrated in Figure 3. A selection of the mid three seconds of the five-second analog voice recordings that were used for acoustical analysis were also used for perceptual judgement ratings. These included nine recorded samples of the vowel /a/ from the last week of baseline and nine recorded samples from the last week of treatment for each of the subjects. A table of random numbers was used to generate nine pairs of voice samples consisting of an end-of-baseline and an end-of-treatment sample. These were spliced together to form a master listening tape for each subject's data. Each pair of voice samples was presented to the listeners with one second between the components of a vowel pair, and three seconds between each pair. The order in which the end-of-baseline and end- of -treatment members of a pair of voice samples occurred on the tape was varied randomly to preclude an order effect on listeners' perceptions of a difference between the paired samples.

The perceptual ratings of each subject's voice were provided by three professional speech-language pathologists in separate listening sessions. All three judges had completed previous course-work in voice disorders and their clinical experience with voice ranged

from one to five years. Each listener sat in a sound-treated room 50 cm in front of a loudspeaker (Fostex 6301 B) to allow optimal presentation ($0^{\circ}$ azimuth) of the paired voice

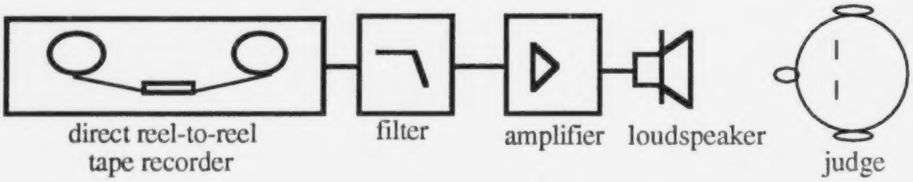


Figure 3. Instrumental array for perceptual data acquisition.

samples. On presentation to the listeners, the recorded signals were low-pass filtered at 8 kHz to condition them comparably to their preparation for digital analysis. Each judge adjusted the volume on the speaker to a comfortable listening level. A free-field presentation was chosen because Wilde, Martens, Hillenbrand and Jones (1986) found that synthetically jittered voice signals presented through earphones sounded less rough than the same signals presented free-field.

The perceptual ratings were made utilizing a visual analogue scale (VAS) for hoarseness adapted by the primary investigator from reports of its use in recent literature (Bond & Lader, 1974; Lomas, Pickard, Bester, Elbard, Finlayson & Zoghaib, 1989). The VAS used in this research is illustrated in Figure 4. Listeners were first informed regarding the proper use of the VAS (Appendix G). One end of the continuum represented the direction toward normal voice quality and the other end, the direction toward dysphonic voice. The components of each pair of voice samples were rated relative to each other on a single VAS. Listeners were requested to make a judgement about the quality of the first vowel of each pair and mark somewhere on the VAS with a "1". The voice quality of the second vowel of the pair was to be marked on the VAS with a "2", relative to the first judgement. The relative positions of the two ratings on the VAS with respect to the ends of the scale allowed the investigator to determine whether a judge perceived each end-of-treatment voice sample as better or worse than the corresponding end-of-baseline sample. The primary investigator was also interested possibly in documenting some evidence of a perceived magnitude difference between the end-of-baseline and end-of-treatment voice sample ratings among judges and across subjects. After perceptual judgements were made, a template marked with millimetre divisions was laid over the VAS to determine approximate magnitude differences in perceived dysphonia. The millimetre difference between rating scores from the VAS were paired with the corresponding jitter difference scores for each vowel pair. These scores provided the data base for the perceptual-acoustic data correlation proposed among the experimental questions.

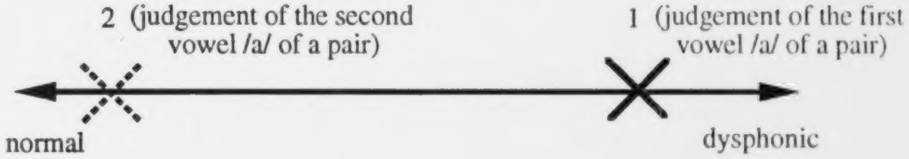

2 (judgement of the second vowel /a/ of a pair)
1 (judgement of the first vowel /a/ of a pair)
normal
dysphonic

Figure 4. Visual analogue scale for perceptual ratings of dysphonia exemplifying a hypothetical case in which the second vowel of a pair was judged as considerably less dysphonic than the first vowel heard.

## Reliability

Acoustical data. Total measurement reliability was obtained by re-digitizing 25% of the recorded voice samples selected randomly from baseline and treatment sessions. Re-editing of the re-digitized signal was considered acceptable if the investigator's measurements fell within one pitch period of the original editing sites that had isolated the mid three seconds of each voice signal for analysis. The investigator's repeated measurements fell within one pitch period 100% of the time. Re-edited waveforms were re-analyzed for all acoustical parameters (jitter, shimmer, $F_0$ and SNR measures). Total measurement accuracy for baseline measures was found to be 100% for subjects three and four. The baseline data accuracy for subjects one and two ranged from 0.001 to 0.006 ms for jitter, 0.01 to 0.14 % for shimmer, 0.01 to 0.60 Hz for $F_0$, and 0.01 to 0.40 dB for SNR. Total measurement accuracy for treatment measures was found to be 100% for the data of subjects two and three. The treatment data of subjects one and four for jitter were 100% reliable; shimmer accuracy ranged from 0.01 to 0.06%; $F_0$ values ranged from 0.01 to 0.40 Hz, and SNR values ranged from 0.01 to 0.05 dB.

Perceptual data. To establish intrajudge reliability on the perceptual ratings, the principal investigator replayed all four master listening tapes for each judge following a ten-minute break. Point-to-point percentage reliability scores were calculated for the two sets of perceptual ratings made by each judge. Judgements were considered in agreement when the repeated rating of the same vowel pair placed each member of the pair in the same position relative to the other with respect to the ends of the VAS. Intrajudge reliability for repeated ratings is tabulated in Table 2.

Table 2. Intrajudge reliability based on overall % agreement for each judge's repeated ratings of vowel pairs on master tape of each subject.

| | Subject 1 | Subject 2 | Subject 3 | Subject 4 |
|---|---|---|---|---|
| Judge 1 | 89% | 56% | 89% | 89% |
| Judge 2 | 89% | 100% | 89% | 100% |
| Judge 3 | 67% | 78% | 100% | 78% |

Interjudge reliability on the perceptual ratings was determined by calculating pair-by-pair percentage agreement scores among the listeners' VAS ratings for the first presentation of the master tape. Judges' ratings were considered in agreement when the three judges' ratings for the same vowel pair placed each member of the pair in the same position relative to the other with respect to the ends of the VAS. Interjudge reliability is tabulated in Table 3.

Table 3. Interjudge reliability, based on % agreement among all three judges' ratings of vowel pairs on each subject's master tape.

| | |
|---|---|
| Subject 1 | 56% |
| Subject 2 | 45% |
| Subject 3 | 89% |
| Subject 4 | 78% |

# CHAPTER IV. RESULTS

## FFVN

A post-treatment FFVN laryngeal examination performed by a certified speech-language pathologist at the University was used to document any structural change in the subjects' vocal nodules following treatment. An otolaryngologist also viewed the pre- and post-treatment FFVN videotapes for each of the subjects and confirmed the following:

Prior to treatment subject one's bilateral vocal nodules were 2-3 mm in size with the nodule on the right vocal fold appearing slightly larger than the one on the left vocal fold. At the end of treatment subject one presented with bilateral vocal nodules that were noticeably reduced in size. Subject two presented with 2-3mm bilateral vocal nodules prior to treatment. Remnants of one vocal nodule on the right vocal fold remained at the end of treatment; the nodule on the left vocal fold had disappeared. Subject three's bilateral vocal nodules were 3-4 mm before treatment. At the end of treatment there was no sign of vocal nodules; slight scarring on the left vocal fold was noted. Subject four exhibited bilateral pinpoint vocal nodules with scarred tissue on the right vocal fold prior to treatment. Both nodules had disappeared by the end of treatment; scarred tissue on the right vocal fold remained. Table 4 provides a summary of the pre- and post-treatment status of each subject's vocal nodules as well as any change in the acoustical data. Documentation of the change in the acoustical data was based on a comparison of the values obtained from the last week of baseline and the last week of treatment. These results will be discussed below.

Table 4. Summary of subject status pre- and post-treatment.

| Subject | Vocal nodules (pre-treatment) | Vocal nodules (post-treatment) | Jitter | Shimmer | $F_0$ | SNR |
|---|---|---|---|---|---|---|
| 1 | bilateral 2-3mm | reduced | ↑ | ↑ | ↑ | ↑ |
| 2 | bilateral 2-3mm | remnants on R vocal fold | ↑ | ↑ | ↓ | ↓ |
| 3 | bilateral 3-4mm | absent | ↑ | ↓ | ↑ | ↑ |
| 4 | bilateral pinpoint | absent | ↔ | ↑ | ↑ | ↓ |

## Menstrual cycle

Instances of the onset of menses are plotted on the graphs of the acoustical data illustrated in Figures 5-8. The menstrual cycle did not appear to influence the acoustic variables across the baseline and treatment phases. This conclusion was based on the fact that no systematic pattern of variation appeared to exist as a function of menstrual cycles for the jitter, shimmer, $F_0$ and SNR magnitudes plotted over time for subjects one, two and three. Subject four did not have a menstrual cycle.

## Acoustical Parameters

### Jitter

Visual analyses of graphs of average jitter were used to determine whether jitter magnitude changed in the data obtained for the sustained productions of /a/ before and during therapy. Figure 5 (A - D) show average jitter values in milliseconds for the vowel /a/ across time (baseline, and treatment) for each of the four subjects. Visual analysis of the data across the treatment phase of the experiment inspected the levels, day-to-day and within-session variability, and overall trends of the data. The average jitter values for sessions in the last week of baseline and the last week of treatment are indicated on the graphs for each of the subjects. These sessions' data were also used for the perceptual judgement ratings and the correlations among the acoustical and

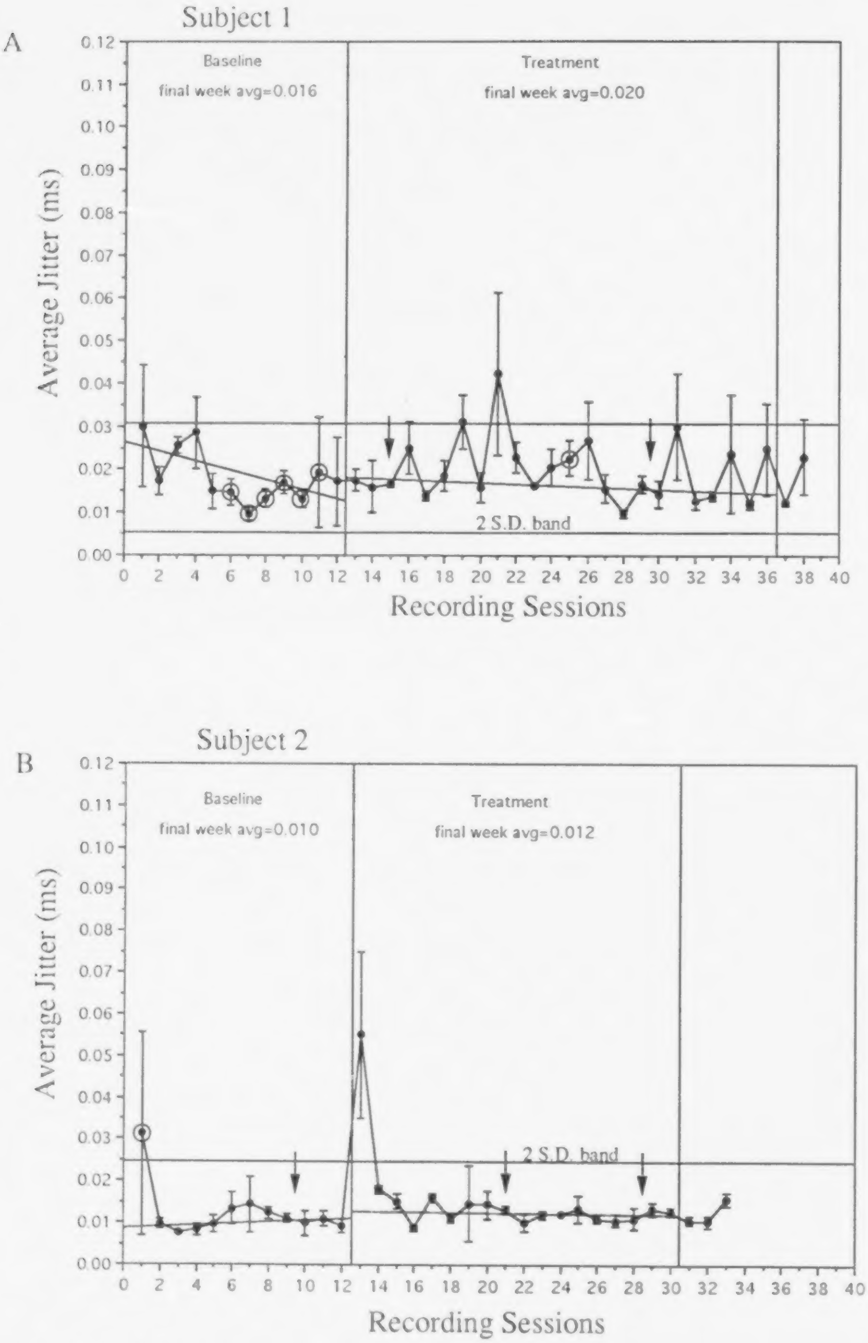


Figure 5. A & B. Average session jitter values in milliseconds plotted across baseline and treatment phases for subjects 1 and 2. The "final week averages" represent the jitter averages for the last three sessions of each phase. Instances of the onset of menses are represented by arrows; sessions when signs of colds/allergies were noticeable are circled. 1 S.D. error bars are indicated by ⫯.

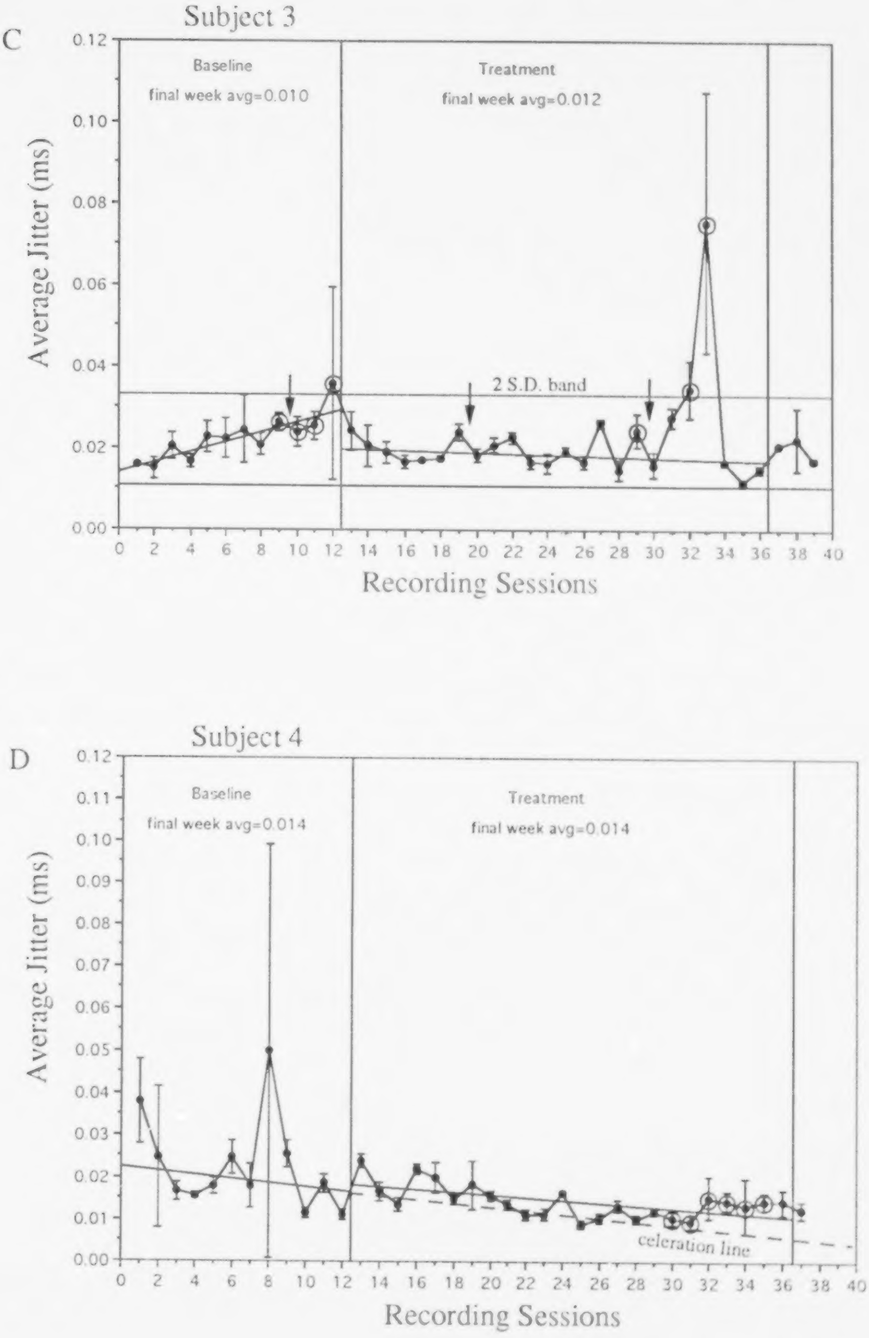


Figure 5. C & D. Average session jitter values in milliseconds plotted across baseline and treatment phases for subjects 3 and 4. The "final week averages" represent the jitter averages for the last three sessions of each phase. Instances of the onset of menses are represented by arrows; sessions when signs of colds/allergies were noticeable are circled. 1 S.D. error bars are indicated by ↕.

perceptual data for each of the subjects. Two-standard-deviation bandwidths based on the baseline data were calculated for visual inspection of the treatment data of subjects one, two and three. This method was chosen because the variable nature of these subjects' data would not allow for unbiased visual inspection using the celeration line approach. A celeration line was computed for the baseline phase of subject four and extended into her treatment phase data to assess change in the acoustical variables. This method was chosen for subject four because her data exhibited serial dependency for jitter, shimmer and fundamental frequency, and the celeration line takes into account serial dependency (Ottenbacher, 1986).

Average jitter values plotted for subjects one through four show slight decreasing trends in jitter for subjects one, three and four across treatment, and stable average jitter levels for subject two with treatment. With the exception of the pre-treatment period for subject two, a stable jitter baseline was not obtained, and a treatment effect could not be demonstrated for any of the subjects. Based on comparison of the average session jitter values from the last week of baseline and the last week of treatment, subject one's jitter levels exhibited a slight rise from 0.016 ms to 0.020 ms, and subject two's average session jitter values rose slightly from 0.010 ms to 0.012 ms. Subject one's day-to-day and within-session jitter values remained variable over the baseline and treatment phases. In contrast, Subject two's day-to-day and within-session jitter values remained stable over the baseline and treatment phases with the exception of two outliers. In contrast, subject three's jitter levels dropped from an end-of-baseline average of 0.029 ms to 0.015 ms near the end of the treatment phase. Her day-to-day jitter values remained variable across the treatment period. This variability may partially be attributed to subject three's poor health, that is, sore throats and frequent colds. Subject four's average jitter values were the same (0.014 ms) for the last week of baseline and the last week of treatment, though her day-to-day and within-session jitter values appeared less variable during the treatment phase.

In summary, based on comparison of the average session jitter values from the last week of baseline and the last week of treatment, three of the four subjects showed a slight increase in average session jitter at the end of treatment, while no change in this variable was observed in the remaining subject. A stable baseline was obtained only for subject two. Only subject four's day-to-day and within-session jitter values appeared less variable with treatment. The data plotted for subjects one, two and three did not exceed a two standard deviation bandwidth preventing any conclusions regarding a treatment effect. Although more data points are represented above the celeration line for subject four, the instability of her baseline prevents any conclusions regarding a treatment effect using the celeration line approach. A celeration line for an unstable baseline biases the reader's interpretation of the results; in the case of subject four's data the declining line suggests that jitter values were higher with treatment, when in fact average jitter values dropped with treatment (but not at the rate that jitter values dropped across baseline measures).

While the focus of this research concerned jitter, average session shimmer, Fo, and SNR values were also plotted in a similar manner and are illustrated in Figures 6-8.

### Shimmer

Average shimmer values plotted for subjects one through four in Figure 6 (A-D) show increasing trends in average shimmer for subjects one and three, a decreasing trend in subject four across treatment, and stable shimmer values for subject two with treatment. Shimmer values behaved in a manner similar to the jitter values for each of the four subjects. Based on comparisons of the average shimmer values from the last week of baseline and the last week of treatment, subject one's shimmer increased slightly from 2.34 % in baseline to 2.50 % at the end of the treatment phase, and large day-to-day and within-session variability remained noticeable across the baseline and treatment phases. Subject two's shimmer values showed little within-session and day-to-day variability over the baseline and treatment phases; shimmer increased slightly from 0.939 % at the end of baseline to 1.2 % at the end of treatment. Subject three's day-to-day shimmer values remained noticeably variable across the experimental period, though her within-session variability for shimmer appeared to be reduced with the onset of treatment. Her shimmer decreased from 3.30 % at the end of baseline to 1.77 % at the end of treatment. Subject four's shimmer values showed less day-to-day and within-session variability during the treatment phase; vocal shimmer increased slightly from 1.71 % at the end of baseline to 1.78 % at the end of treatment.

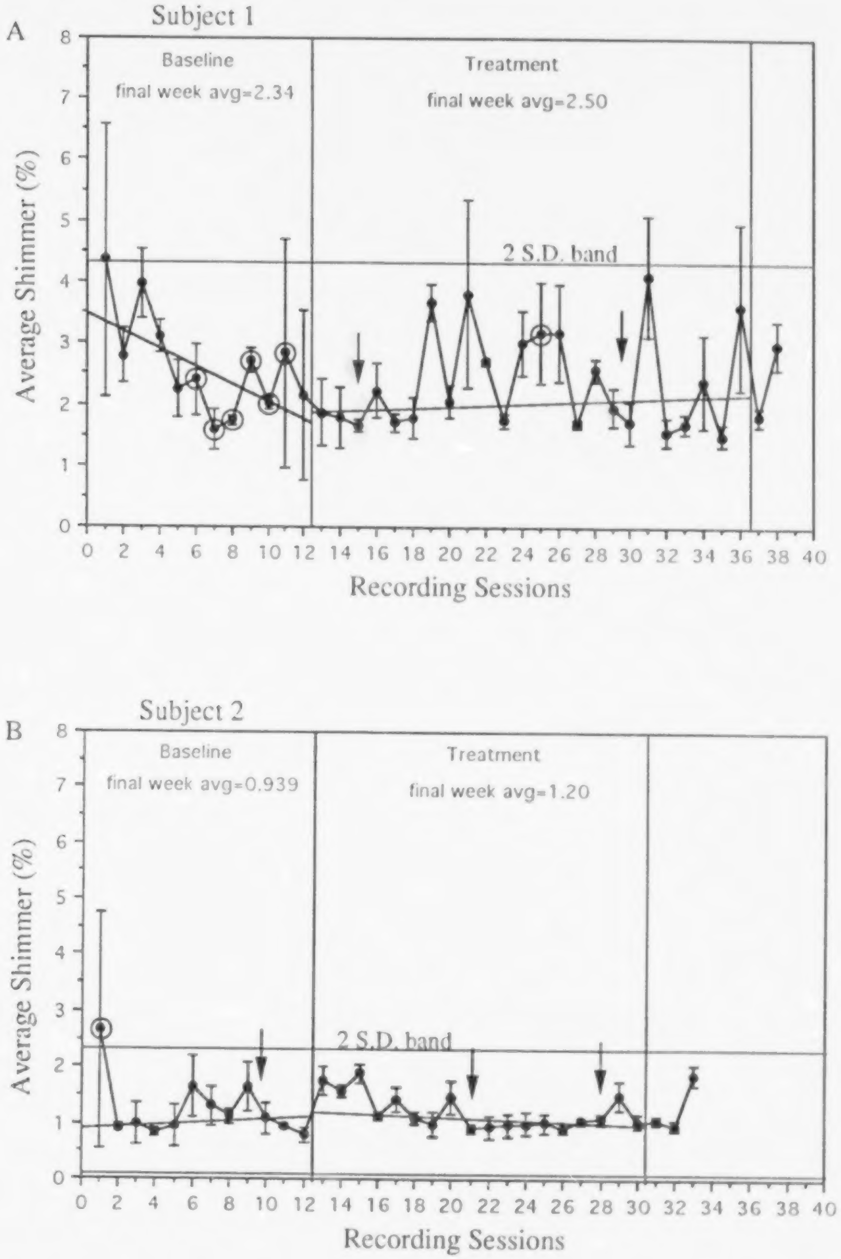


Figure 6. A & B. Average session shimmer values in percent plotted across baseline and treatment phases for subjects 1 and 2. The "final week averages" represent the shimmer averages for the last three sessions of each phase. Instances of the onset of menses are represented by arrows; sessions when signs of colds/allergies were noticeable are circled. 1 S.D. error bars are indicated by ⫯ .

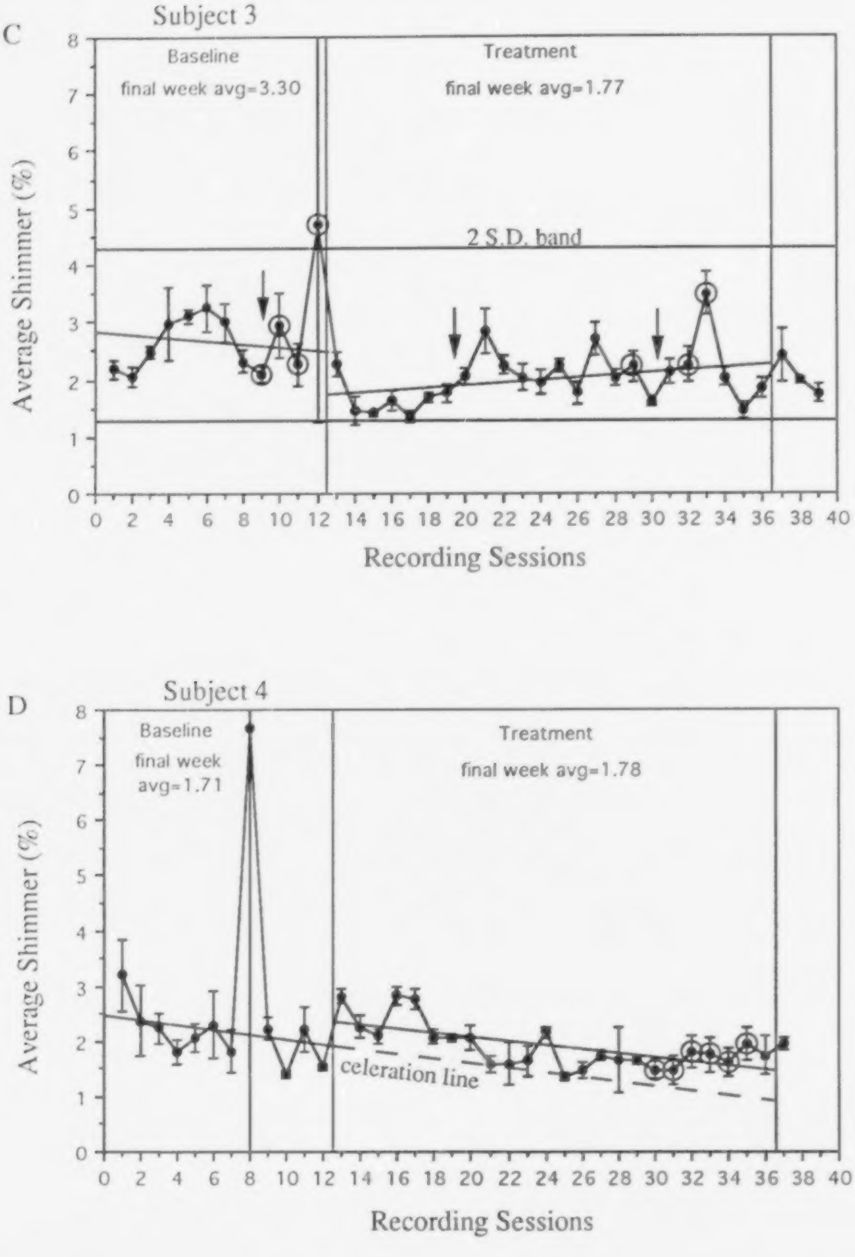


Figure 6. C & D. Average session shimmer values in percent plotted across baseline and treatment phases for subjects 3 and 4. The "final week averages" represent the shimmer averages for the last three sessions of each phase. Instances of the onset of menses are represented by arrows; sessions when signs of colds/allergies were noticeable are circled. 1 S.D error bars are indicated by $\overset{\top}{\underset{\bot}{\bullet}}$.

### Fundamental frequency

Average fundamental frequency values plotted for subjects one through four in Figure 7 (A-D) show increasing trends in Fo for subjects one, three and four and a stable Fo for subject two with treatment. A stable baseline (stationary trend) was obtained only for subject one. Based on comparison of the average fundamental frequency values obtained from the last week of baseline and the last week of treatment, subject one's average session Fo values rose from 198.0 to 225.6 Hz with treatment. Subject two's average Fo dropped from 207.5 to 186.5 Hz with treatment, and the variability in her day-to-day Fo values appears to have decreased with treatment. Subject three's average Fo increased from 153.9 at the end of baseline to 179.4 at the end of treatment, and a change in trend from falling to rising Fo, was noted with treatment. Subject four's Fo values increased from 210.97 at the end of baseline to 230.9 Hz at the end of treatment. Within-session average fundamental frequency values showed minimal variability across the baseline and treatment phases in all four subjects. Between-session average fundamental frequency values for all subjects remained variable across treatment sessions. Fundamental frequency data plotted for subjects two and three did not exceed a two standard deviation bandwidth; subject one's data exceeded the two standard deviation bandwidth indicating a treatment effect (Ottenbacher, 1986). Although more data points are represented below the celeration line for subject four, the instability of her baseline prevents any conclusions regarding a treatment effect using the celeration line approach. A celeration line for an unstable baseline biases the reader's interpretation of the results; in the case of subject four's data the ascending line suggests that fundamental frequency values were lowered with treatment, when in fact average fundamental frequency values increased with treatment.

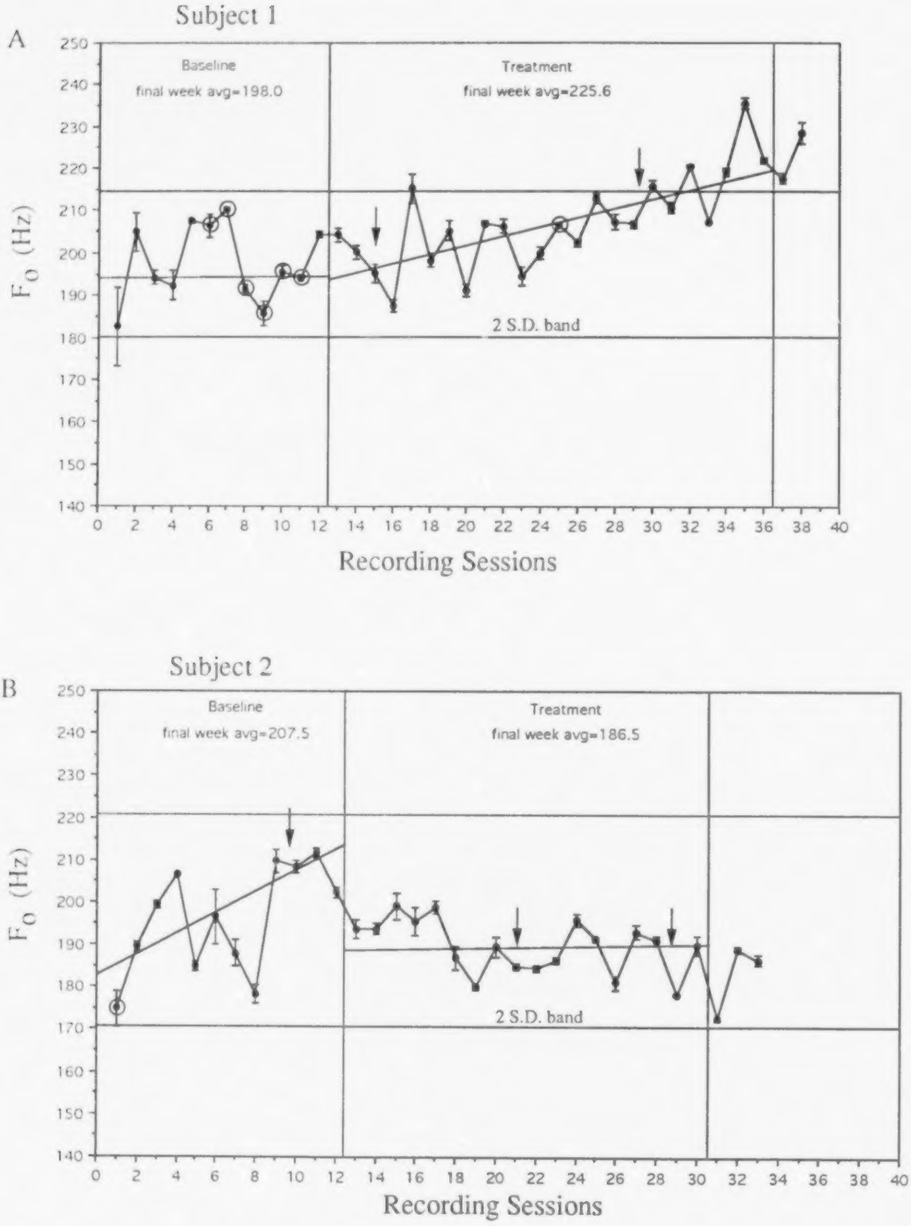


Figure 7. A & B Average session $F_0$ values in Hertz plotted across baseline and treatment phases for subjects 1 and 2. The "final week averages" represent the $F_0$ averages for the last three sessions of each phase. Instances of the onset of menses are represented by arrows; sessions when signs of colds/allergies were noticeable are circled. 1 S.D. error bars are indicated by I .

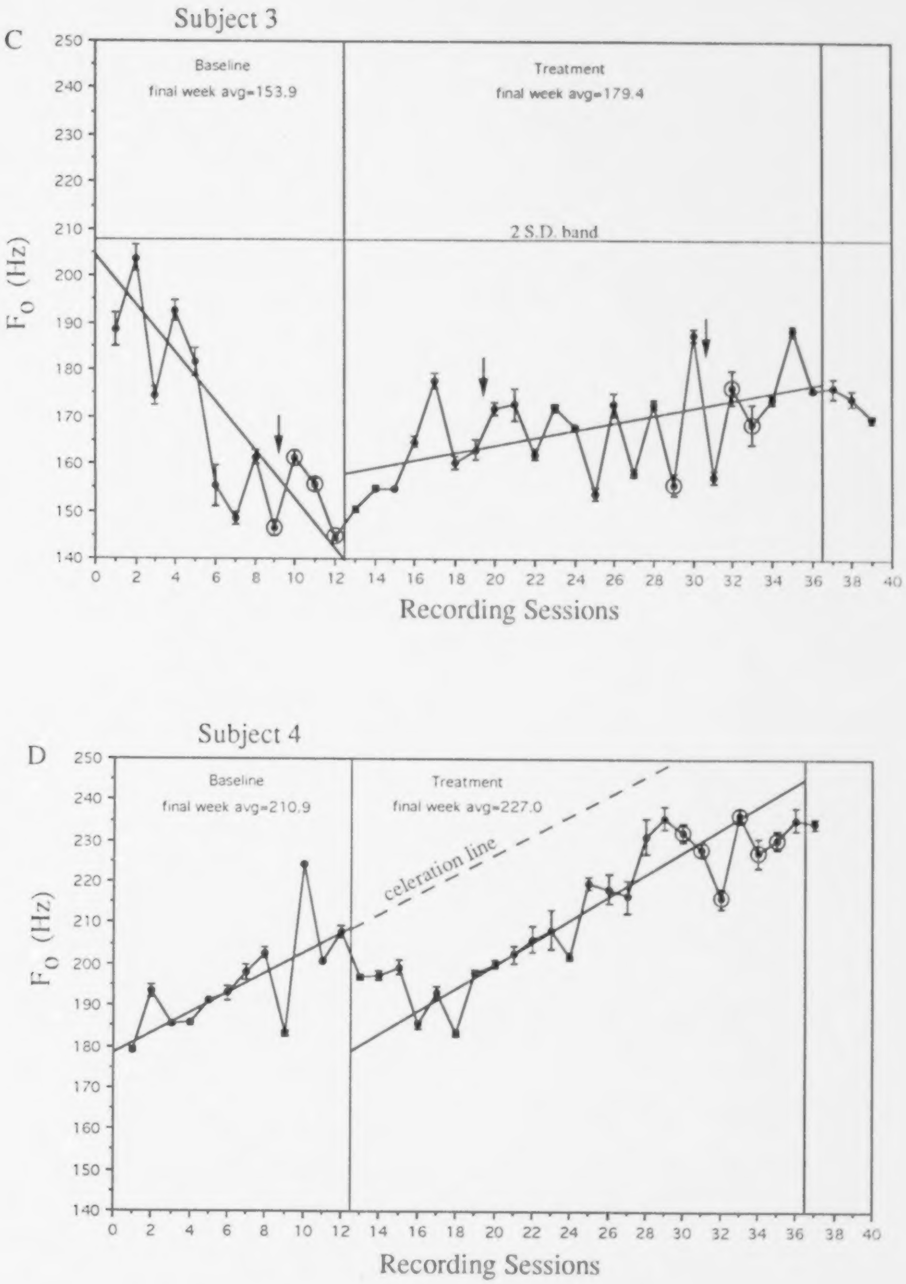


Figure 7. C & D Average session $F_0$ values in Hertz plotted across baseline and treatment phases for subjects 1 and 2. The "final week averages" represent the $F_0$ averages for the last three sessions of each phase. Instances of the onset of menses are represented by arrows; sessions when signs of colds/allergies were noticeable are circled. 1 S.D. error bars are indicated by ⫯.

### Signal-to-noise ratio

Average SNR's plotted for subjects one through four in Figure 8 (A-D) show increasing trends for subjects one, two and four, and remained stationary in subject three across treatment. However, based on comparisons of the average SNR values obtained from the last week of baseline and the last week of treatment, average session SNR values rose only slightly with treatment for subjects one and three, and decreased slightly in subjects two and four. Subject one's SNR rose from 24.51 dB at the end of baseline to 25.29 dB at the end of treatment. Subjects three's average SNR rose from 20.32 to 24.82 dB. Subject two's data showed a slight drop from 29.59 in baseline to 29.07 dB with treatment. Subject four's SNR decreased slightly from an end-of-baseline level of 26.26 to 26.20 dB at the end of treatment. Between-session and within-session average SNR remained variable across the treatment phases in all four subjects. The SNR data plotted for subjects one and two did not exceed a two standard deviation bandwidth; subject three's data exceeded the two standard deviation bandwidth indicating a treatment effect (Ottenbacher, 1986). Although more data points are represented below the celeration line for subject four, the instability of her baseline prevents any conclusions regarding a treatment effect using the celeration line approach. A celeration line for an unstable baseline biases the reader's interpretation of the results; in the case of subject four's data the ascending line suggests that SNR values were much lower with treatment, when in fact average SNR values decreased only slightly with treatment.

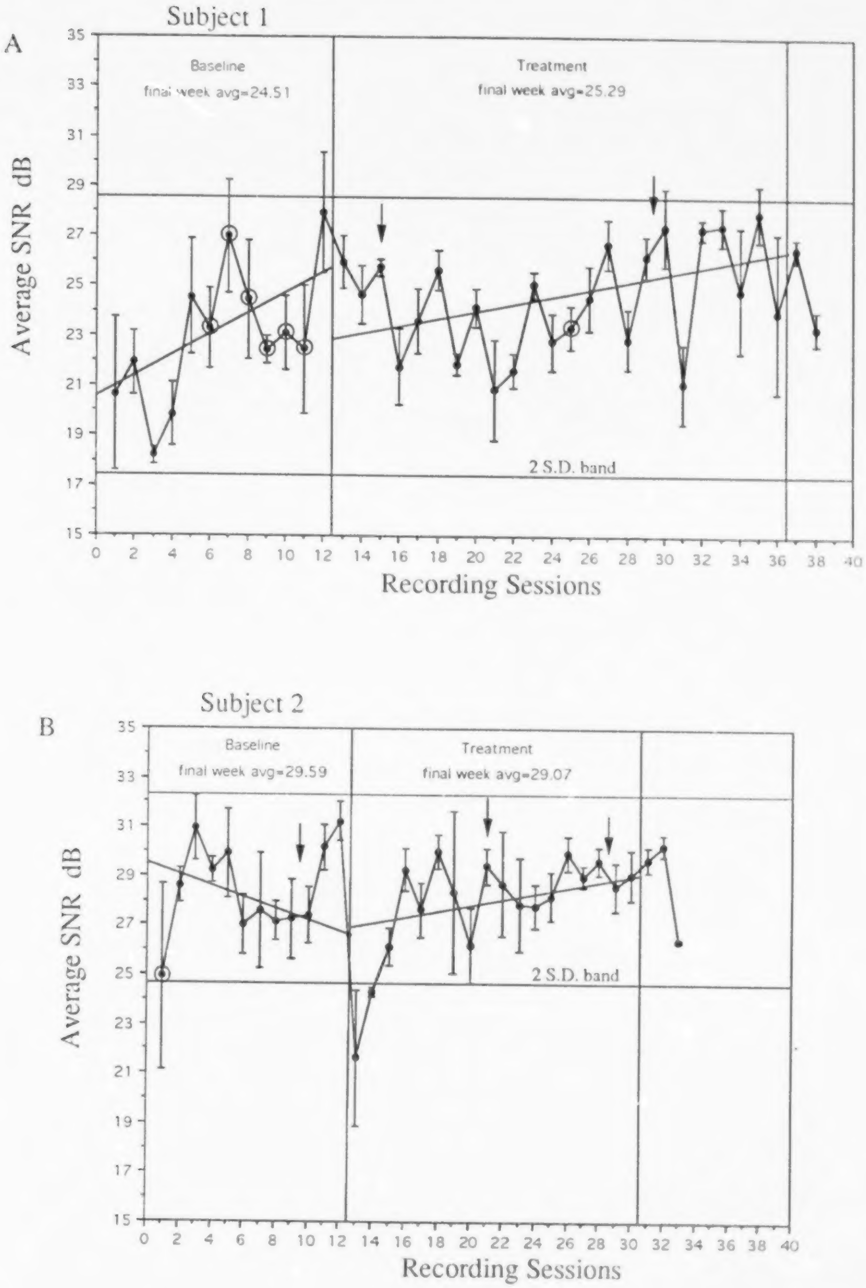
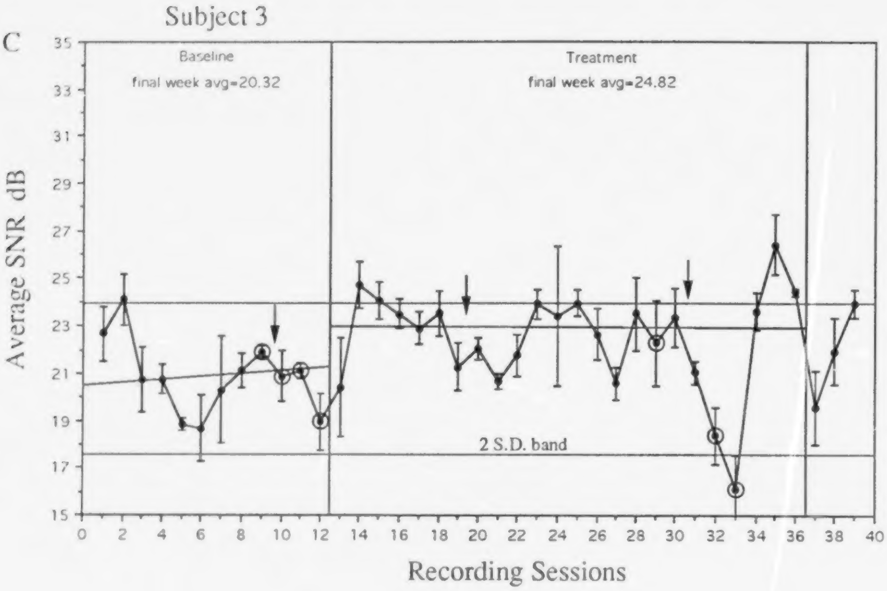


Figure 8. A & B Average session SNR values in decibels plotted across baseline and treatment phases for subjects 1 and 2. The "final week averages" represent the SNR averages for the last three sessions of each phase. Instances of the onset of menses are represented by arrows; sessions when signs of colds/allergies were noticeable are circled. 1 S.D. error bars are indicated by Ī.



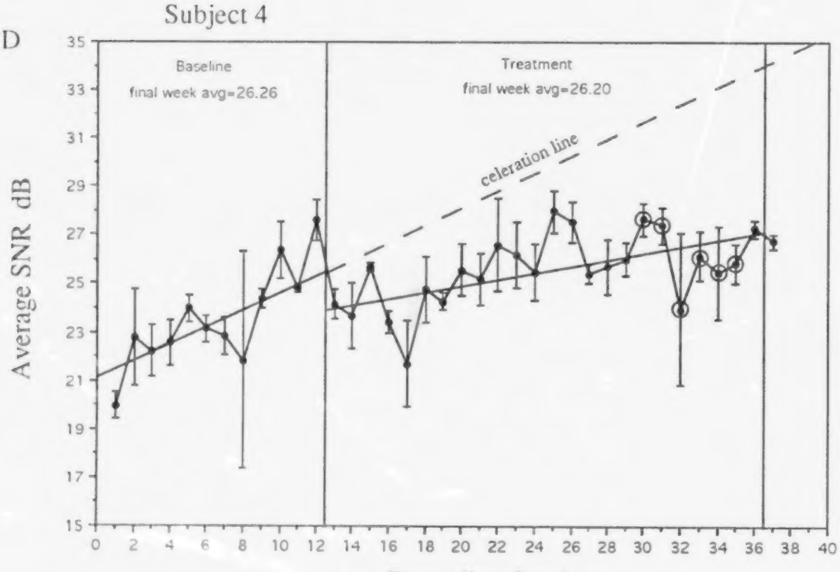


Figure 8. C & D Average session SNR values in decibels plotted across baseline and treatment phases for subjects 3 and 4. The "final week averages" represent the SNR averages for the last three sessions of each phase. Instances of the onset of menses are represented by arrows; sessions when signs of colds/allergies were noticeable are circled. 1 S.D. error bars are indicated by ⧗ .

## Multiple Correlations

Table 5 shows a summary of the average jitter, shimmer, SNR and $F_0$ difference values obtained for the paired samples of /a/ from the last week of baseline and the last week of treatment. The difference values were calculated by subtracting the end-of-treatment values from the end-of-baseline values for each of the acoustic variables. This resulted in a difference value with a negative sign if the end-of-treatment values were larger than those obtained at the end-of-baseline. Thus, a jitter difference value with a negative sign indicates that jitter increased at the end-of-treatment; a jitter difference value with a positive sign indicates that jitter decreased with treatment. These average data values were used for multiple correlations among the acoustical variables.

Multiple comparisons using Pearson product-moment correlations (Bruning & Kintz, 1977) were used to determine the relation among the average difference values for jitter, shimmer, $F_0$ and SNR for the vowel /a/ from the last week of baseline and the last week of therapy. The per-experiment error rate was set at 0.05 for each set of six comparisons, and the per-comparison error rate was 0.01 (Kirk, 1982). Table 6 illustrates the multiple correlations among these variables. In all subjects, negative correlations were found between jitter and fundamental frequency. Positive correlations were found between jitter and shimmer values. Negative correlations were found between jitter and SNR. Shimmer and fundamental frequency, and shimmer and SNR showed negative correlations. Of the four correlations between $F_0$ and SNR, three were positive, and one was negative. Only three of the correlations were significant at the .01 level; these were between jitter and shimmer for subjects one ($r= .91$), three ($r= .926$) and four ($r= .827$) respectively. Scattergrams representing these multiple correlations are illustrated in Figure 9 (A-C). The scattergrams provide a visual guide for the reader to aid in the interpretation of these correlations. For example, the positive correlations between jitter and shimmer differences for subjects one through four are plotted in

| Subject | Baseline session. | Treatment session. | /a/ | Baseline Jitter | Treatment Jitter | Difference Jitter (B-T) | Baseline Shimmer | Treatment Shimmer | Difference Shimmer (B-T) | Baseline $F_o$ | Treatment $F_o$ | Difference $F_o$ (B-T) | Baseline SNR | Treatment SNR | Difference SNR (B-T) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 11 | 24 | 1 | .011 | .013 | -.002 | 1.94 | 2.06 | -.120 | 194.59 | 221.35 | -26.76 | 23.24 | 27.62 | -4.38 |
| 1 | 11 | 24 | 2 | .034 | .027 | .007 | 4.97 | 4.25 | .720 | 193.84 | 221.70 | -27.86 | 19.61 | 22.60 | -2.99 |
| 1 | 11 | 24 | 3 | .013 | .034 | -.021 | 1.58 | 4.52 | -2.940 | 194.34 | 222.54 | -28.20 | 24.55 | 21.67 | 2.88 |
| 1 | 12 | 23 | 1 | .009 | .011 | -.002 | 1.13 | 1.57 | -.440 | 203.57 | 234.23 | -30.66 | 29.69 | 29.03 | .66 |
| 1 | 12 | 23 | 2 | .014 | .013 | .001 | 1.60 | 1.64 | -.040 | 204.28 | 236.31 | -32.03 | 28.93 | 26.81 | 2.12 |
| 1 | 12 | 23 | 3 | .029 | .012 | .017 | 3.74 | 1.30 | 2.440 | 205.20 | 236.14 | -30.94 | 25.17 | 28.02 | -2.85 |
| 1 | 10 | 22 | 1 | .015 | .022 | -.007 | 2.14 | 2.23 | -.090 | 194.69 | 218.94 | -24.25 | 22.06 | 25.21 | -3.15 |
| 1 | 10 | 22 | 2 | .014 | .038 | -.024 | 2.00 | 3.22 | -1.220 | 197.27 | 218.51 | -21.24 | 22.52 | 22.20 | .32 |
| 1 | 10 | 22 | 3 | .011 | .011 | .000 | 1.92 | 1.73 | .190 | 194.56 | 220.10 | -25.54 | 24.83 | 27.21 | -2.38 |
| 2 | 10 | 18 | 1 | .013 | .012 | .001 | 1.09 | 1.13 | -.040 | 209.32 | 192.45 | 16.87 | 26.50 | 29.44 | -2.94 |
| 2 | 10 | 18 | 2 | .010 | .012 | -.002 | 1.35 | .88 | .470 | 206.74 | 188.60 | 18.14 | 26.98 | 29.80 | -2.82 |
| 2 | 10 | 18 | 3 | .007 | .014 | -.007 | .80 | 1.07 | -.270 | 209.51 | 188.63 | 20.88 | 28.67 | 27.88 | .79 |
| 2 | 12 | 16 | 1 | .008 | .010 | -.002 | .92 | 1.11 | -.190 | 202.70 | 192.24 | 10.46 | 31.37 | 29.36 | 2.01 |
| 2 | 12 | 16 | 2 | .011 | .014 | -.003 | .78 | 1.16 | -.380 | 203.55 | 190.25 | 13.30 | 30.38 | 29.18 | 1.20 |
| 2 | 12 | 16 | 3 | .009 | .009 | .000 | .66 | .97 | -.310 | 201.08 | 191.02 | 10.06 | 31.93 | 30.26 | 1.67 |
| 2 | 11 | 17 | 1 | .012 | .013 | -.001 | .93 | 1.71 | -.780 | 210.39 | 178.43 | 31.96 | 29.40 | 27.81 | 1.59 |
| 2 | 11 | 17 | 2 | .009 | .012 | -.003 | .92 | 1.22 | -.300 | 212.38 | 178.73 | 33.65 | 31.18 | 29.72 | 1.46 |
| 2 | 11 | 17 | 3 | .012 | .015 | -.003 | 1.00 | 1.60 | -.600 | 212.35 | 177.93 | 34.42 | 29.90 | 28.21 | 1.69 |
| 3 | 12 | 22 | 1 | .063 | .016 | .047 | 8.68 | 1.90 | 6.780 | 145.15 | 172.92 | -27.77 | 17.60 | 24.12 | -6.52 |
| 3 | 12 | 22 | 2 | .019 | .017 | .002 | 2.40 | 2.20 | .200 | 143.69 | 173.74 | -30.05 | 19.91 | 22.73 | -2.82 |
| 3 | 12 | 22 | 3 | .026 | .017 | .009 | 3.04 | 1.96 | 1.080 | 145.43 | 174.82 | -29.39 | 19.39 | 24.03 | -4.64 |
| 3 | 10 | 24 | 1 | .021 | .014 | .007 | 2.36 | 1.81 | .550 | 163.09 | 175.67 | -12.58 | 21.81 | 24.45 | -2.64 |
| 3 | 10 | 24 | 2 | .023 | .015 | .008 | 2.95 | 2.03 | .920 | 159.72 | 176.17 | -16.45 | 21.14 | 24.56 | -3.42 |
| 3 | 10 | 24 | 3 | .028 | .016 | .012 | 3.51 | 1.68 | 1.830 | 161.30 | 175.63 | -14.33 | 19.72 | 24.27 | -4.55 |
| 3 | 11 | 23 | 1 | .029 | .011 | .018 | 1.83 | 1.63 | .200 | 156.67 | 188.99 | -32.32 | 21.45 | 25.33 | -3.88 |
| 3 | 11 | 23 | 2 | .026 | .012 | .014 | 2.48 | 1.37 | 1.110 | 156.07 | 187.43 | -31.36 | 20.85 | 26.11 | -5.26 |
| 3 | 11 | 23 | 3 | .022 | .012 | .010 | 2.47 | 1.38 | 1.090 | 154.25 | 189.50 | -35.25 | 21.03 | 27.80 | -6.77 |
| 4 | 12 | 22 | 1 | .012 | .021 | -.009 | 1.47 | 1.84 | -.370 | 206.55 | 223.10 | -16.55 | 26.72 | 23.29 | 3.43 |
| 4 | 12 | 22 | 2 | .010 | .009 | .001 | 1.53 | 1.32 | .210 | 207.71 | 229.46 | -21.75 | 28.35 | 26.78 | 1.57 |
| 4 | 12 | 22 | 3 | .012 | .010 | .002 | 1.59 | 1.69 | -.100 | 209.49 | 228.85 | -19.36 | 27.76 | 26.33 | 1.43 |
| 4 | 10 | 24 | 1 | .011 | .015 | -.004 | 1.47 | 1.76 | -.290 | 223.98 | 232.32 | -8.34 | 27.03 | 27.66 | -.63 |
| 4 | 10 | 24 | 2 | .013 | .011 | .002 | 1.37 | 1.40 | -.030 | 224.13 | 236.11 | -11.98 | 25.03 | 27.21 | -2.18 |
| 4 | 10 | 24 | 3 | .011 | .017 | -.006 | 1.36 | 2.08 | -.720 | 224.38 | 237.67 | -13.29 | 27.03 | 26.93 | .10 |
| 4 | 11 | 23 | 1 | .016 | .015 | .001 | 1.74 | 2.16 | -.420 | 200.62 | 227.73 | -27.11 | 24.65 | 25.61 | -.96 |
| 4 | 11 | 23 | 2 | .020 | .012 | .008 | 2.45 | 1.62 | .830 | 200.92 | 231.15 | -30.23 | 24.89 | 26.74 | -1.85 |
| 4 | 11 | 23 | 3 | .020 | .016 | .004 | 2.45 | 2.09 | .360 | 200.92 | 232.10 | -31.18 | 24.89 | 25.20 | -.31 |

Table 5. Summary of average baseline, treatment and average difference values (end-of-baseline minus end-of-treatment) for jitter (col 7), shimmer (col 10), $F_o$ (col 13) and SNR (col 16) for all subjects. The order of baseline and end-of-treatment session numbers reflect the way they were originally randomly paired.

Table 6. Correlations among average difference magnitudes of jitter, shimmer, $F_0$ and SNR in paired samples of sustained /a/ obtained during the final week of baseline and the final week of treatment. (* = significant at the 0.01 level)

| Subject | Variables | Correlation | R-squared |
|---|---|---|---|
| 1 | jitter & $F_0$ | -.615 | .378 |
| 2 | | -.228 | .052 |
| 3 | | -.173 | .03 |
| 4 | | -.68 | .462 |
| | | | |
| 1 | jitter & shimmer | .91* | .829 |
| 2 | | .09 | .008 |
| 3 | | .926* | .858 |
| 4 | | .827* | .683 |
| | | | |
| 1 | jitter & SNR | -.523 | .273 |
| 2 | | -.313 | .098 |
| 3 | | -.602 | .362 |
| 4 | | -.604 | .365 |
| | | | |
| 1 | shimmer & $F_0$ | -.315 | .099 |
| 2 | | -.47 | .221 |
| 3 | | -.022 | .0004 |
| 4 | | -.638 | .407 |
| | | | |
| 1 | shimmer & SNR | -.652 | .424 |
| 2 | | -.768 | .59 |
| 3 | | -.611 | .374 |
| 4 | | -.314 | .098 |
| | | | |
| 1 | $F_0$ & SNR | -.26 | .068 |
| 2 | | .192 | .037 |
| 3 | | .52 | .27 |
| 4 | | .127 | .016 |

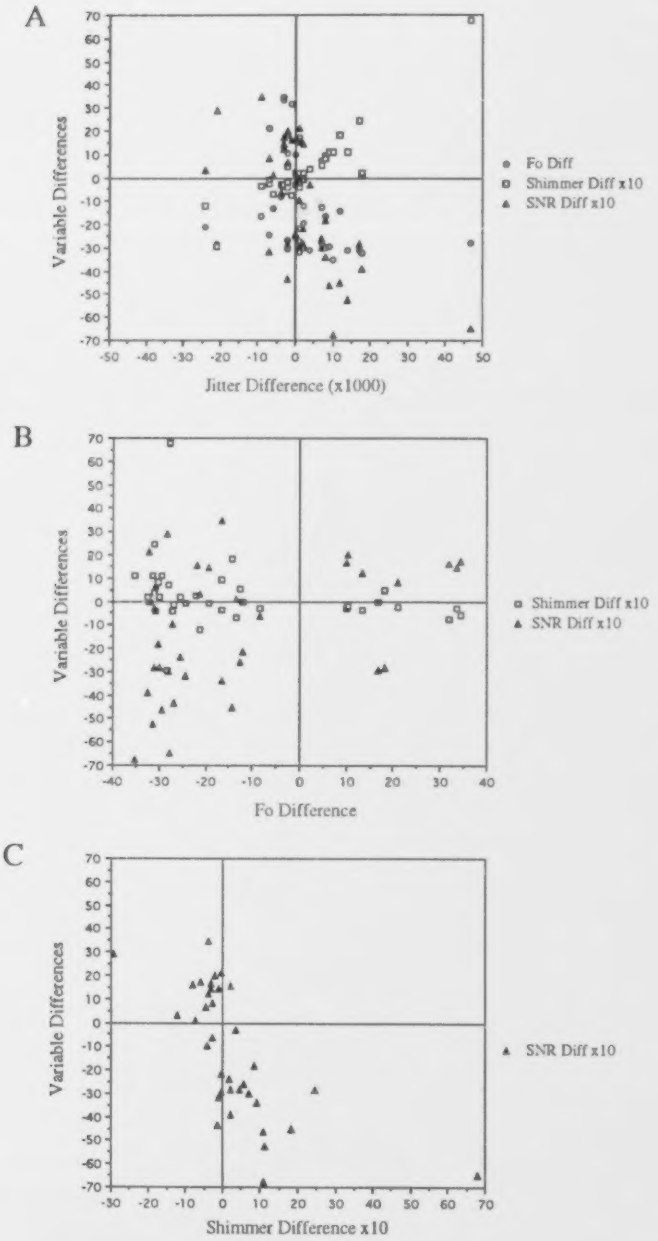


Figure 9. Scattergrams illustrating multiple correlations among average difference values (end-of-baseline minus end-of-treatment) for jitter, shimmer, $F_0$ and SNR for all subjects.

Figure 9A. The squares representing shimmer for the most part follow a positive slope representing a positive relationship with jitter with respect to the behavior of these two variables across treatment. The squares that deviate most from the straight line are those of subject two. The negative correlations between jitter and $F_0$ are represented by the circles that approximate a negative slope in Figure 9A. The negative correlations between jitter and SNR are represented by the triangles that approximate a negative slope in that figure.

The negative correlations between shimmer and $F_0$, and the mostly positive correlations between SNR and $F_0$ for subjects one through four are plotted in Figure 9B. The squares representing shimmer that approximate a slightly negative slope represent its equivocal relationship with $F_0$. The positive correlations between SNR and $F_0$ for subjects two, three and four are illustrated by the triangles that approximate a positive slope in Figure 9B. The negative correlation between SNR and $F_0$ for subject one is represented by the triangles that approximate a negative slope in Figure 9B.

The correlations between shimmer and SNR for subjects one through four are plotted in Figure 9C. The negative correlations between shimmer and SNR are represented by the triangles that approximate a negative slope.

## Perceptual ratings

### Reliability

All judges perceived a difference in the subjects' voices between paired samples of sustained /a/ obtained during the final week of baseline and the final week of treatment. Table 2 (p. 53) shows the intrajudge reliability ratings for the voice samples for each subject. Overall intrajudge reliability scores were best for subject three (89, 89 and 100% ) and poorest for subject two (56, 100 and 78%). Table 3 (p. 54) shows the interjudge reliability ratings for the voice samples for each subject. Judges exhibited the worst consensus rating subject two (45%) and achieved the best consensus in their ratings for subject three (89%). Figure 10 shows the number of times the judges rated the end-of-treatment voice samples as better (toward the normal end of the VAS) or worse (toward the dysphonic end) than the comparable baseline sample. They rated the end-of-treatment voice samples of subjects one, three and four as about the same or better than the end-of-baseline samples. Overall the judges rated the quality of subject two's end-of-treatment voice samples as the same as, or worse than, that of her baseline samples.

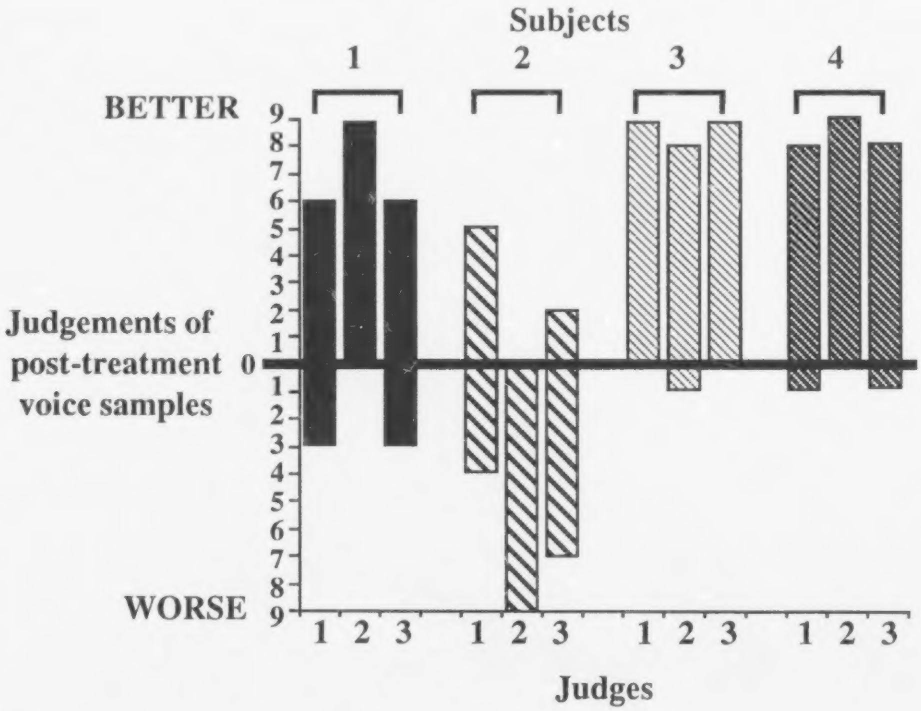


Figure 10. Perceptual ratings of end-of-treatment voice samples relative to end-of-baseline sample ratings.

### Correlations with jitter

The primary purpose in using the listeners' perceptual ratings of the recorded voice samples was to document any directional change in subjects' voice qualities (either better or worse) following voice treatment for vocal nodules. The primary investigator also attempted to investigate the subjective magnitude of any perceived change by applying a millimetre template to the VAS following the listeners' ratings.

Pearson product-moment correlations (Bruning & Kintz, 1977) were used to determine the degree to which differences in jitter magnitudes (in ms) correlated with perceived magnitude differences in voice quality (in VAS mm) for paired voice samples from the last week of baseline and the last week of therapy. Table 7 illustrates the correlations between the jitter magnitude difference values for pairs of sustained /a/ obtained during the last week of baseline and the last week of therapy and the perceptual rating difference scores for the same pairs of /a/ samples for all subjects. [The average jitter end-of-treatment value was subtracted from the average jitter end-of-baseline value for each of the paired vowels. The perceptual difference rating scores were obtained by measuring the difference in millimetres on the VAS between ratings of the same paired samples of /a/ from the last week of baseline and the last week of treatment]. Predominantly positive correlations were found between jitter magnitude difference values and perceived differences in dysphonia. Judge one's perceptual difference ratings demonstrated significant positive correlations with jitter difference magnitudes for subjects one through three at the 0.05 level. Judge two's ratings of perceived voice quality difference demonstrated mostly positive correlations with jitter difference magnitudes for all subjects, although none of the correlations was significant at the 0.05 level. For judge three, positive correlations between jitter and perceptual difference ratings were significant only for subjects one and four. There does not appear to be one

Table 7. Jitter magnitude difference values between paired samples of sustained /a/ obtained during the final week of baseline and the final week of treatment correlated with the perceived difference in dysphonia between the same pairs of /a/ samples in all four subjects. (* = significant at the 0.05 level)

| Subject | Judge | Count | Covariance | Correlation | R-squared |
|---|---|---|---|---|---|
| 1 | one | 9 | 0.301 | 0.697* | 0.486 |
| 2 | one | 9 | 0.018 | 0.631* | 0.399 |
| 3 | one | 9 | 0.233 | 0.738* | 0.545 |
| 4 | one | 9 | 0.016 | 0.187 | 0.035 |
| | | | | | |
| 1 | two | 9 | 0.114 | 0.532 | 0.284 |
| 2 | two | 9 | 0.016 | 0.599 | 0.359 |
| 3 | two | 9 | -0.748 | -0.912 | 0.831 |
| 4 | two | 9 | 0.021 | 0.312 | 0.098 |
| | | | | | |
| 1 | three | 9 | 0.463 | 0.692* | 0.479 |
| 2 | three | 9 | -0.023 | -0.287 | 0.083 |
| 3 | three | 9 | 0.041 | 0.272 | 0.074 |
| 4 | three | 9 | 0.13 | 0.814* | 0.662 |

subject whose jitter difference magnitudes demonstrated better correlations with the perceptual difference ratings provided by each of the judges.

Scattergrams illustrating each judge's perceived difference score derived from the position of her ratings of vowel pairs on the VAS are plotted against the jitter difference values for the same vowel pairs for all subjects in Figure 11 (A-C). These scattergrams illustrate the fact that although the relative jitter difference values did not correlate significantly with the relative difference values in perceived dysphonia for each of the subjects, many of the judges' ratings were consistent with changes in jitter at the end of treatment. The diagonal lines shown on the scattergrams for judges one, two and three indicate the slope along which the dots representing each subject's data should fall for a positive correlation between jitter difference values and perceptual ratings. Therefore, dots in the upper right-hand quadrant indicate instances when judges rated the end-of-treatment voices as better, and jitter values decreased at the end of treatment.

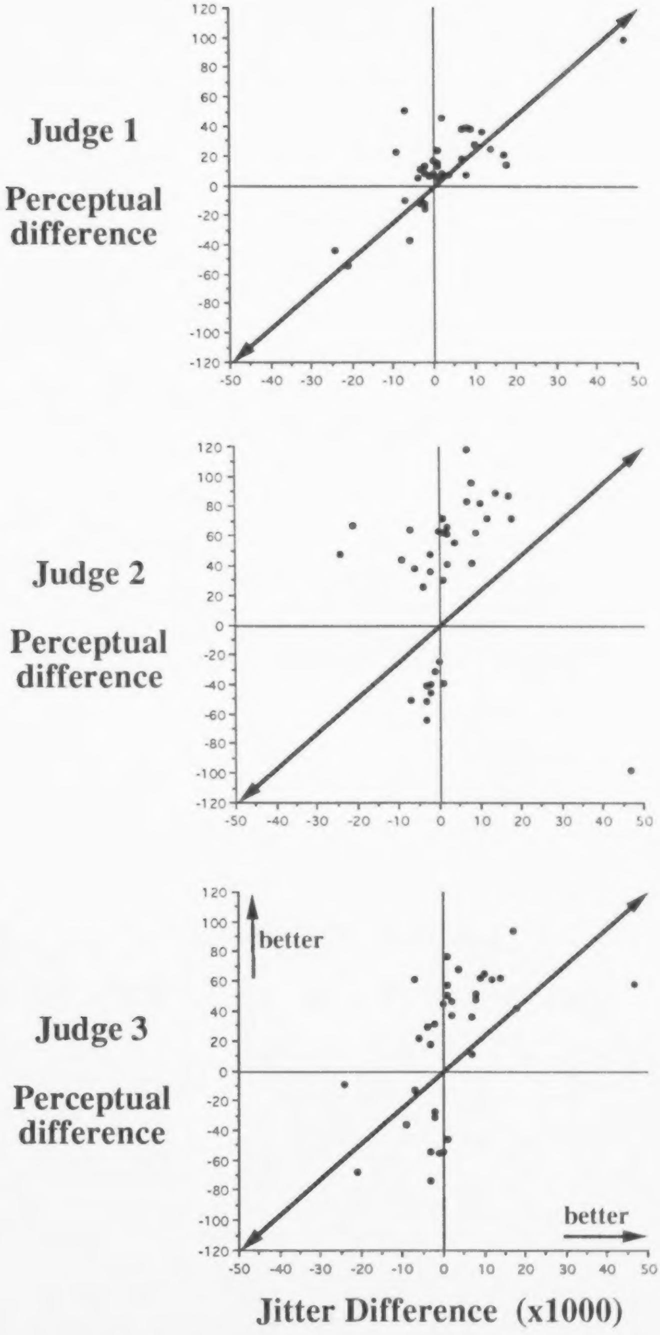


Figure 11. Scattergrams illustrating each judge's perceived difference scores derived from ratings of vowel pairs on the VAS plotted against the jitter difference values for the same vowel pairs for all subjects. Each dot represents the average jitter difference value plotted against the perceived difference rating for a single vowel pair.

The dots in the lower left-hand quadrant represent instances in which judges rated the voices as worse, and jitter values increased with treatment. The dots shown in the upper left-hand quadrant indicate instances in which the judges rated the voices as better, when in fact, jitter magnitudes had increased slightly following treatment. Finally, dots in the lower right-hand quadrant represent instances in which the voice was judged to be worse, when in fact, jitter values had decreased following treatment. This result occurred in only one instance. From the scattergrams it also appears that those paired samples on which judges had more difficulty rating a change in dysphonia were ones in which changes in jitter magnitudes were minimal. End-of-treatment voice samples rated toward the normal rather than the dysphonic end of the VAS scale by all three judges relative to the baseline sample included cases in which jitter difference values had increased slightly.

The issue of what variable, or combination of acoustic variables, contributed significantly to the perception of change in voice quality originally proposed among the experimental questions in this research was beyond the scope of this study. A multiple regression prediction equation could not be used to determine the relative contributions of each of the four acoustical variables (jitter, shimmer, $F_0$ and SNR) to the perception of change in voice quality over time due to the limited number of data points. This question remains an important one for future research.

# CHAPTER V. DISCUSSION

The primary purpose of this research was to compare acoustical and perceptual methods for monitoring changes in the dysphonic voice with voice therapy. Jitter, shimmer, $F_0$ and SNR were measured in the sustained /a/ productions of four adult females with vocal nodules before and during voice therapy. These same vowel samples from the last week of baseline and the last week of treatment were used for perceptual judgement ratings. The relationship between jitter measures and listener's perceptions of dysphonia was also investigated. The results of this study will be discussed relative to the five research questions posed. Clinical implications, limitations of the present study as well as suggestions for future research will also be reviewed.

It is first important to discuss several issues that pertain to the interpretation of the acoustical and perceptual data. These issues concern the influence of the menstrual cycle on jitter, shimmer, $F_0$ and SNR levels, the availability of a stable baseline against which treatment effects may be compared and the difficulty in comparing the acoustical results with other studies of vocal parameters.

## Menstrual cycle

Similar to the outcomes of studies reporting no noticeable differences in acoustical variables prior to or with the onset of menstruation, results from the present research indicate no obvious relationship between fluctuations in the acoustical data and the onset of menses during the baseline and treatment phases. This conclusion was based on the fact that jitter and shimmer magnitudes shown in Figures 5 & 6 (pp. 57-58 & 62-63) demonstrated fluctuations across the baseline and treatment phases that did not appear to coincide with any particular stage of the menstrual cycle. No systematic pattern of variation existed for the jitter, shimmer, $F_0$ and SNR magnitudes plotted over time for subjects one, two and three. Subject four did not have a menstrual cycle. These results must be interpreted with caution however due to the small sample size (n=3) and the possible influence of other factors (ie. colds, psychological and physical stress and an individual subject's ability to control perturbation levels) that may have influenced fluctuations of the data. In addition, the present study did not investigate the influence of the time of ovulation associated with changes in jitter levels. Higgins and Saxman (1989) measured jitter in the sustained vowels /i/ and /u/ transduced by a throat contact microphone and analyzed via computer analysis. Recordings were collected every other day over a 33-day sampling period for 10 women (aged 19 - 25 years) and in 5 men who acted as a control group. Perturbation values for the vowels /i/ and /u/ were combined as no significant differences between vowels for women were found. Results for the women were analyzed in relation to fluctuations of ovarian hormones inferred from basal body temperature data and subjects' reports of the onset of menstruation. Mean jitter values for each subject were computed and plotted over time. No similarity in the pattern of variation over time was seen between groups or subjects. They found no significant differences in the average vocal jitter magnitudes of women at premenstruation and the onset of menstruation. A change greater than 1 SD from the subject's mean perturbation value was considered to be significant. They did however discover notable upward and downward changes in jitter magnitudes at the time of ovulation for most subjects. Specifically when standardized jitter factor values from ovulation and postovulation were pooled to test the significance of jitter around the time of ovulation, change in jitter around the time of ovulation was significantly greater than that which would have been predicted based on chance alone. Higgins and Saxman (1989) speculate that findings of increased variation in jitter at ovulation lends support to the idea that when ovarian hormones fluctuate the vibratory stability of the vocal folds is affected. Further they state that these influences may be strong enough to influence the voices of non-singers during conditions of normal speaking.

It is possible that increased control for variables known to effect jitter and more precise monitoring of the effects of ovulation on a larger sample may show changes in the acoustical variables that were not evident in the present study.

### Baseline Stability

The present research employed an A-B single-subject design to allow for pretreatment tracking of the acoustic variables and documentation of their change across a period of treatment. This design allows one to observe the variability within subjects and to replicate treatment effects across subjects. Evaluation of treatment effects within an individual is made possible by visual comparison of the data points in the A and B phases of the study. In comparing these phases one must consider the stability of the baseline, variability within and between phases, changes in the trends and level of the data, and as well the overall evaluation of the patterns of the data (McReynolds & Kearns, 1983).

It is important to establish that the behavior is not changing before treatment is initiated in order to evaluate change associated with the effects of treatment. The ideal is to establish a stable baseline and then compare these data to those which are obtained from the treatment phases. This is of course an ideal and unusual event when one is dealing with human subjects. With respect to the present study, baselines established for jitter and shimmer were variable with the exception of those for subject two. Subject two and four's $F_0$ and SNR baselines were also variable.

The fact that stable baselines could not be obtained for each subject limits the interpretation regarding a treatment effect and widens the range that the treatment data must exceed in order to demonstrate any significant change. Jitter, shimmer, $F_0$ and

SNR data plotted in Figures 5-8 for each of the subjects, indicate that the data had become less variable during the last week of baseline. For this reason, the primary investigator decided to analyze changes in the acoustic data based on a comparison of the jitter, shimmer, $F_0$ and SNR values that were obtained from the last week of baseline and the last week of treatment. These same data points also provided the data base for the acoustical/perceptual comparisons.

## Comparing the results of different studies

The third issue that remains to be considered concerns the need for caution in comparing results obtained from different studies that have used widely different sampling and analysis techniques (Horii, 1980).

Jitter and shimmer "normative" values may differ from one study to the next depending on the methods used in marking the individual periods of vocal fold movement, the sampling rate and duration, the perturbation formulae used in perturbation calculations, and varying controls for variables known to effect jitter including speaker sex, age, fundamental frequency, and intensity. In addition, it is difficult to compare values across subjects because there is quite often an overlap of values reported for normal and pathological voices.

Table 5 (p. 71) shows the end-of baseline jitter values reported for each of the subjects in the present study. These range from 0.009 ms (9 µs) up to 0.063 ms (63 µs), and the end-of-treatment jitter values range from 0.009 ms (9 µs) to 0.038 ms (38 µs). These values are as low or lower than the range of 16 to 89 µs reported by Horii (1979) for normal speakers. He measured jitter at a sampling rate of 40kHz using a peak-picking $F_0$ analysis program (SEARP). The end-of treatment jitter values are closer to the mean jitter values of 0.015 ms reported for women with no signs of vocal pathology by Milenkovic (1987), the author of the jitter analysis program employed in the present study. Karnell et al. (1991) compared the results reported for jitter and shimmer data measured in the sustained vowel /a/ productions of one man and one woman. These authors reported that jitter and shimmer measurement obtained using the Visi-Pitch were not comparable with those obtained using two different minicomputers and suggest that standardization among recording and analysis procedures is required before data from different laboratories can be compared.

The only study to date that has applied acoustic analysis to tape-recorded voice samples from 41 patients before and after successful voice therapy was that of Askenfelt and Hammarberg (1986), but they did not report raw vocal jitter and shimmer values for each of the patients before and after therapy. Because other studies in the literature have not tracked jitter and shimmer values across treatment in voice-disordered subjects there are no similar studies for comparison. With these points in mind results of the present study will only be considered relative to each other (pre-treatment and end-of-treatment values) for each of the acoustic variables across the four subjects.

Research question number one asked whether a difference in jitter magnitude levels existed for the mid portions of sustained productions of /a/ before and during voice therapy in the four adult females with vocal nodules who served as the subjects in this experiment.

### Jitter

The results obtained from these four subjects reveal that jitter for the mid-portion of the sustained vowel /a/ does not show much absolute change or change in a consistent direction with treatment. In addition, the relative change in jitter magnitudes does not appear to vary in a consistent manner with changes in the size of a subject's vocal nodules (see Table 4. p. 56).

Graphic analysis and visual inspection of the jitter magnitude levels plotted in Figure 5 (A-D; pp. 57-58) for the mid-portions of sustained productions of /a/ before and during voice therapy did not reveal a consistent change in jitter for subjects one through four across treatment. Based on a comparison of the average jitter levels from the last week of baseline and the last week of treatment, it appears that the average session jitter increased slightly in three of the subjects at the end of treatment, and showed no change in the remaining subject. The average jitter values plotted for each of the subjects in Figure 5 (A-D; pp. 57-58) show slight decreasing trends in jitter for subjects one, three and four across treatment, and stable jitter values for subject two during treatment.

It is interesting to consider these results in the context of changes in subjects' vocal nodules with treatment. The absolute change in jitter values did not appear consistent with the change in the size of the vocal nodules noted via FFVN at the end of treatment. Based on a comparison of the average jitter values from the last week of baseline and the last week of treatment, subject one's session jitter levels indicated a slight rise ( 0.004 ms), when in fact her nodules appeared to decrease in size, from 2-3 mm at the outset of the experiment to pinpoint nodes on FFVN examination following treatment.

Subject two's average session jitter rose slightly (0.002 ms) during the last week of treatment, although one of her bilateral nodules, sized at 2-3mm before treatment, resolved altogether, and the other was just a remnant at the end of treatment. Nevertheless, her day-to-day and within-session jitter values remained relatively stable over the baseline and treatment phases. This may have been due to the fact that she is a trained singer and might therefore tend to exhibit less day-to-day and within-session variability in her jitter values. Murry and Large (1979) suggest that voice training may allow singers to exhibit extraordinary control over the expiratory airstream during voicing thus allowing them to decrease their perturbation magnitudes.

Subject three's jitter levels dropped (0.034 ms) with treatment. Her 3-4mm nodules disappeared completely although slight scarring was noted on the left vocal fold at the end of treatment. Subject three's poor health interferes with the accurate interpretation of her data. Overall, average session jitter values may have dropped more noticeably across treatment if illness (colds) had not elevated her jitter values on sessions 9, 10, 11 and 12 of baseline, and sessions 17, 20 and 21 of the treatment phase. It is interesting to note that subject three's average jitter levels demonstrated the most absolute change with treatment, and this coincided with the fact that her nodules exhibited the largest change in size among all the subjects.

Subjects four's average jitter values were the same (0.014 ms) for the last week of baseline and the last week of treatment even though her vocal nodules, which were admittedly the smallest before therapy, had disappeared by the end of treatment. Subject four's average session jitter values may have shown a larger decrease if she had not been afflicted with cold-like symptoms brought on by allergies during the last week of treatment.

While the focus of this research concerned jitter, differences in average session shimmer, $F_0$ and SNR across treatment were examined in a similar manner as these variables are known to covary with jitter (Hillenbrand, 1987; Milenkovic, 1987).

## Shimmer

The results obtained from these four subjects suggest that shimmer for mid-portions of sustained /a/, like jitter values in that context, did not show much absolute change or vary in a consistent direction with treatment. With the exception of subject two's data, shimmer values remained variable over the baseline and treatment phases. In addition, similar to the results obtained for jitter, shimmer magnitudes did not appear to vary in a manner consistent with the change in size of a subject's vocal nodules (see Table 4 p. 56).

Graphic analysis and visual inspection of the shimmer session values across baseline and treatment phases revealed that shimmer behaved in a manner similar to that of the jitter values for each of the four subjects. This result is not unexpected because jitter and shimmer are known to covary (Hillenbrand, 1987; 1988). Based on comparisons of the average shimmer values from the last week of baseline and the last week of treatment, it appears that shimmer increased slightly in three of the four subjects with treatment, and decreased with treatment in one. The average session shimmer values plotted in Figure 6 (A-D; pp. 62-63) show trends of increasing average shimmer in two of the subjects and a decreasing trend or stable pattern for the other two subjects. Only subject four's shimmer data deviate slightly from the trends expected based on jitter. While her jitter values remained unchanged at the end of treatment, her shimmer values increased slightly.

Similar to the patterns for jitter, absolute change in shimmer values did not appear consistent with the change in the size of the vocal nodules noted via FFVN at the end of treatment. Based on a comparison of the average shimmer values from the last week of baseline and the last week of treatment, subject one's shimmer levels indicated a slight rise (0.16%), when in fact her nodules appeared to decrease in size, from 2-3mm at the outset of the experiment to pinpoint nodes on FFVN examination following treatment.

Subject two's shimmer increased slightly (0.261%) during the last week of treatment, although one of her bilateral vocal nodules sized at 2-3 mm prior to treatment resolved altogether, and the other was just a remnant at the end of treatment. Her shimmer values showed little within-session and day-to-day variability over the baseline and treatment phases.

Subject three's vocal shimmer dropped (1.53 %) with treatment. Her 3-4mm nodules disappeared completely, although day-to-day shimmer values remained noticeably variable across the experimental period. Within-session variability for shimmer appeared to be reduced with the onset of treatment. Subject three's high day-to-day variability may have been elevated due to illness (colds) on sessions 9, 10, 11 and 12 of baseline, and sessions 17, 20 and 21 of treatment. Her shimmer values might have shown even more absolute change during treatment if colds had not been an influencing factor.

Subject four's shimmer values increased slightly (0.07 %) at the end of treatment, although her pinpoint nodules disappeared completely with treatment. Her shimmer data showed less day-to-day and within-session variability during the treatment phase. As with this subject's jitter values, her shimmer values may have shown a larger decrease if she had not been afflicted with allergies during the last week of treatment.

### Fundamental frequency

Average fundamental frequency values plotted for subjects one through four in Figure 7 (A-D; pp. 65-66) show increasing trends in $F_0$ for subjects one, three and four and a stable $F_0$ for subject two with treatment. Based on comparison of the average fundamental frequency values obtained from the last week of baseline and the last week of treatment, average session $F_0$ values increased in subjects one (27 Hz), three (25.5 Hz) and four (20 Hz), and dropped by 19 Hz in subject two with treatment. Similar to the values obtained for jitter and shimmer these $F_0$ results may be interpreted in the context of changes in subject's vocal nodules with treatment. Unlike the results obtained for jitter and shimmer, however, changes in the direction of fundamental frequency appear to be more consistent with the change in the size of the nodules for each of the subjects following treatment (see Table 4 p. 56). Research has shown that subjects with a laryngeal growth such as vocal nodules which alters the mass of the vocal folds exhibit a reduced $F_0$ (Murry & Doherty, 1980; Davis 1981; Boone, 1983). If this cause-effect relationship is true, it is reasonable to expect that as the mass on the vocal folds decreases, average fundamental frequency would increase.

Based on comparison of the values obtained from the last week of baseline and the last week of treatment, average $F_0$ rose in subjects one, three and four. These values coincided with a noticeable reduction in the size of the nodules in subject one, and complete disappearance of the nodules in subjects three and four. Once again subject three demonstrated the most relative change in her $F_0$ and the most change in her nodules. Subject two's average $F_0$ dropped slightly with treatment, even though her day-to-day $F_0$ values had stabilized and remnants of only one vocal nodule remained on the right vocal fold. Subject two's drop in fundamental frequency is difficult to explain. She frequently complained of excess phlegm in her throat during the baseline and treatment phases but did not appear to suffer from colds during the study. At one point she stated that she has had a very low-pitched speaking voice for several years.

It is important to mention that although raising the fundamental frequency of the subjects' voices was not a direct focus of treatment, $F_0$ may have been influenced by one of the techniques used in the treatment package. Imagery to train the subjects about "placement of the voice in the oral-facial mask" (Boone, 1971) was designed to reduce strain in the laryngeal area and improve vocal quality by achieving a better balance for vocal resonance, but it may also indirectly change pitch to a more natural level. Changes in fundamental frequency may have been influenced by both changes in the mass on the vocal folds and the use of the technique of placing the voice. However, based on the results obtained in the acoustic variables and coincident changes in the vocal nodules for these four subjects it appears that changes in fundamental frequency may provide a better measure of absolute change in vocal pathology and perceptual change in the voice.

### SNR

Based on comparisons of the average SNR values obtained from the last week of baseline and the last week of treatment, it appears that average session SNR values rose slightly in two of the subjects and decreased slightly in the others. Between-and within-session average SNR remained variable across the treatment phases in all four subjects. The average SNR values plotted for each of the subjects in Figure 8 (A-D; pp. 68-69) show increasing trends in SNR for subjects one, two and four, and stabilized SNR values for subject three with treatment. Hillenbrand (1988) suggests that noise levels are higher and SNR's smaller in dysphonic voices than in normal voices; SNR's are correlated with perceived breathiness. The mass-size increases associated with vocal nodules on the medial edges of the vocal folds prevent them from approximating completely. As voice is produced under these circumstances, the outgoing airstream escapes between the poorly approximated margins of the vocal folds and is perceived as a breathy component in the voice (Boone, 1983). One would expect that as the mass of the nodules decreases or disappears and the folds are able to approximate completely the voice will no longer be perceived as breathy. This decrease in vocal pathology and breathy dysphonia should also coincide with an increase in the signal-to-noise ratio. However, as has been mentioned, the degree of "hoarseness" or "breathiness" of the voice is influenced by two factors: 1) the constant leakage of air through a poorly-closed glottis and 2) the period and amplitude fluctuations of the glottal wave (Isshiki et al. 1966; Yanagihara, 1967). As the vocal nodules decrease or disappear, one source of noise (the air leakage) may be reduced, but spectral noise associated with perturbations in the (quasi) periodic signal still remains in the signal. This may partially explain why the absolute change in SNR values tracked across this project did not appear consistent with the corresponding change in the size of the vocal nodules noted via FFVN at the end-of-treatment for each of the subjects (see Table 4, p. 56).

Based on a comparison of the average SNR values from the last week of baseline and the last week of treatment, subject one's SNR levels indicated a slight rise (0.78 dB), and her nodules appeared to decrease from 2-3 mm at the outset of the experiment to pinpoint nodes on FFVN examination at the end of treatment. Subject two's average session SNR decreased (0.52 dB), although one of her bilateral vocal nodules sized at 2-3mm before treatment, resolved altogether, and the other was just a remnant at the end of treatment. Subject three's SNR increased the most (4.5 dB), when compared to SNR changes in the other subjects, and this change coincided with a complete disappearance of her 3-4 mm nodules at the end of treatment. Subject four's SNR decreased slightly (0.06 dB) even though her pinpoint nodules had disappeared at the end of treatment.

In summary, jitter and shimmer measured in the sustained /a/ productions of four adult females did not show much absolute change or change in a consistent direction with voice therapy. Table 4 (p. 56) which shows a summary of each subject's acoustic data along with the status of her vocal pathology following treatment indicates that jitter, shimmer and SNR values did not always change in a direction that would seem consistent with the physical change in each of the subject's nodules. However, fundamental frequency appears to have exhibited the most change with treatment and to have changed in a direction most consistent with the change in the status of each subject's nodules. Fundamental frequency may provide a better index of change in the voice with therapy.

## Multiple Correlations

Research question number two explored the relation among the average difference values for jitter, shimmer, $F_0$ and SNR for the mid-portions of the sustained vowel /a/ from the last week of baseline and the last week of treatment. Multiple comparisons were made among the acoustical data using Pearson product-moment correlations (Bruning & Kintz, 1977).

In interpreting the results of the correlations for the present study one must consider the correlations among these variables that have been reported in the literature. Hillenbrand (1987) found that changes in either jitter, shimmer or SNR measured in synthetic vowels effected changes in all the other acoustic variables. He found that adding increasing amounts of noise to a periodic signal resulted in corresponding significant increases in jitter and shimmer measures. In addition, increasing jitter measures resulted in significant increases in shimmer values. Milenkovic (1987) also found increased shimmer values were associated with increases in jitter values; SNR's decreased with increased jitter in synthetic vowels. Thus, it was not possible to control these parameters independently of one another using synthetically generated voice signals. Observations concerning the relationship between jitter and $F_0$ values have been equivocal (Horii, 1979). Researchers have reported both increasing jitter levels with increases in $F_0$ (Hollien et al. 1973) and increasing jitter levels with decreases in $F_0$ (Jacob, 1968; Horii 1979).

The direction among the multiple correlations obtained for average jitter, shimmer, SNR and $F_0$ difference values in the present study is in general agreement with those obtained from the previously mentioned studies (with the exception of $F_0$ and SNR for subject one). These multiple correlations are illustrated in Table 6 (p. 72). Of all the multiple correlations, only three, between jitter and shimmer, were significant at the 0.01 level. These significant positive correlations indicate that a change in jitter was associated with a change in shimmer values in the same direction. This correlation was not significant for subject two.

Negative correlations were found between jitter and fundamental frequency for all subjects. That is, as jitter changed, fundamental frequency changed in the opposite direction. These data are in agreement with the fact that observations concerning the relationship between $F_0$ level and $F_0$ perturbations are equivocal (Horii, 1979; Orlikoff & Baken, 1990). Negative correlations also were found between jitter and SNR; that is, jitter changes in one direction were associated with changes in SNR in the opposite direction. These results are in agreement with Kempster (1984, cited in Hillenbrand 1987) who also reported negative correlations between jitter and HNR (-0.71) and shimmer and HNR (-0.73). Shimmer and fundamental frequency showed negative correlations; a change in shimmer was associated with a change in $F_0$ in the opposite direction, and shimmer and SNR were negatively correlated; a change in shimmer was associated with a change in SNR's in the opposite direction. In contrast to these results, the data of Deal and Emanuel (1978) showed positive correlations between shimmer and spectral noise levels and jitter and spectral noise levels for vowels sustained by 20 adult males. Finally, fundamental frequency and SNR showed positive correlations except in subject one's data; that is, $F_0$ and SNR's changed in the same direction.

These results must be interpreted with caution however due to the small sample size. The fact that only three of the correlations were significant at the 0.01 level may be partially attributed to the small sample.

Research question number three was concerned with whether listeners heard a difference in the severity of dysphonia between paired samples of sustained /a/ obtained during the last week of baseline and the last week of treatment.

Table 5 (p. 71) shows the nine paired voice samples from the last week of baseline and the last week of treatment that the judges were asked to rate. The order of the baseline and end-of-treatment session numbers reflects the way they were originally randomly paired. Also included are the jitter, shimmer, $F_0$ and SNR difference values between the same vowel pairs for each of the subjects. The difference values for each of the acoustic variables were obtained by subtracting the end-of-treatment values from the end-of-baseline values for each of the paired voice samples.

All judges were able to perceive a difference in severity of dysphonia between paired samples of sustained /a/ obtained during the final week of baseline and the final week of treatment. However, Tables 2 (p. 53) and 3 (p. 54) for intra- and interjudge reliability indicate that the judges were unable to perceive differences among subjects' voice samples with the same degree of consistency. Intra- and interjudge reliability scores were best for the voice of subject three and worse for that of subject two. These ratings are understandable if one considers the difference values for the acoustic parameters of the paired voice samples that the judges were asked to rate. Subject three's difference (baseline minus treatment) values for jitter, shimmer, $F_0$ and SNR are greater than the values shown for the other three subjects. Subject two's jitter, shimmer, $F_0$ and SNR difference values show the least change when compared to those of subject one, three and four. These data suggest that judges may have found it easier to rate subjects' voices that demonstrated relatively more change in their acoustical data with treatment.

The judges' overall ratings of subjects' voices as better or worse following treatment also appear reasonable in view of the corresponding structural changes in the subjects' vocal nodules at the end of treatment. The end-of-treatment voice samples for subjects one, three and four were rated overall as better than their end-of-baseline samples. These ratings coincided with a significant reduction in the size of subject one's nodules, and complete disappearance of the nodules in subjects three and four. In subject two, whose voice was judged overall as the same or worse at the end of treatment, remnants of one vocal nodule were apparent on the right vocal fold, and her acoustic data showed the least change.

All three judges commented on their general difficulty in perceiving change in and reliably rating all of the subjects' end-of-baseline and treatment paired voice samples. However they stated that among the four subject's voice recordings, the most difficult to rate were the voice samples of subject two. When the recorded voice samples of subject three were played back for the listeners they all expressed that her voice was easier to rate than any of the other subjects. They suggested that perhaps if they had had more cues regarding voice onset/offset characteristics, a longer sample which may have included the additional cues of pitch breaks, or a sample of reading or spontaneous speech they may have been better able to rate all of the voices more accurately and reliably. With respect to the length of the voice sample, a three second sample was chosen for analysis for several reasons. The length of the sample used was similar to those that have been reported in the literature. A three second sample also was chosen for practical reasons; the length of time required to digitize and analyze each three-second signal was approximately twenty minutes. In addition, with respect to this research, in order for the acoustic and perceptual data to be comparable for comparative purposes it was necessary to give the judges the exact sample that the computer received for digitization and analysis.

Research question number four was designed to explore the relationship between changes in jitter magnitudes associated with treatment and perceived changes in subjects' voices at the end of treatment. Pearson product-moment correlations were used to determine the degree to which differences in jitter magnitudes (in ms) correlated with perceived differences in voice quality (in VAS mm) for vowel samples paired from the last week of baseline and the last week of therapy. Table 8 shows the end-of-baseline and end-of-treatment voice samples that were paired and spliced together to form a master listening tape for each of the subjects. The jitter difference values were obtained by subtracting the end-of-treatment values from the end-of-baseline values for each of the paired samples of /a/. This resulted in a jitter difference with a negative sign when jitter increased following treatment and a jitter difference with a positive sign when jitter decreased at the end of treatment. The same jitter difference values were paired with the perceptual difference ratings to calculate the acoustical/perceptual correlations for each of the four subjects. Evidence from the literature suggests that increased jitter in pathologic voices is associated with hoarseness and correlates positively with perceptual roughness ratings (Coleman, 1969; Deal & Emanuel, 1978). If increased jitter and perceptual ratings of hoarseness are positively correlated, then a perceptual improvement in the voice (decreased hoarseness) should coincide with a reduction in absolute jitter values following treatment.

The correlations between jitter difference values and perceived dysphonia were positive, but only the correlations between judge three's rating of subject four's voice samples and that subject's jitter difference was significant ($p < 0.05$). It is surprising that the correlations between the acoustical and perceptual data were not consistently significant for any of the subjects. Based on the differences shown in the acoustical data of subject three, for example, as opposed to those for subject two, and the judges' comments regarding their difficulty in rating some voices more than others, it might have been reasonable to expect that the correlations would have been significant for subject three and not for subject two, the trained singer. However, even though the correlations were not significant judges were able to rate the majority of voices in a

| Subject | Baseline session. | Treatment session. | /a/ | Baseline Jitter | Treatment Jitter | Difference Jitter (B-T) | Perceptual Difference Judge #1 | Perceptual Difference Judge #2 | Perceptual Difference Judge #3 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 11 | 24 | 1 | .011 | .013 | -.002 | -16 | 48 | 31 |
| 1 | 11 | 24 | 2 | .034 | .027 | .007 | 18 | 83 | 11 |
| 1 | 11 | 24 | 3 | .013 | .034 | -.021 | -54 | 67 | -68 |
| 1 | 12 | 23 | 1 | .009 | .011 | -.002 | 8 | 36 | -27 |
| 1 | 12 | 23 | 2 | .014 | .013 | .001 | 24 | 72 | 77 |
| 1 | 12 | 23 | 3 | .029 | .012 | .017 | 21 | 87 | 94 |
| 1 | 10 | 22 | 1 | .015 | .022 | -.007 | 51 | 64 | 61 |
| 1 | 10 | 22 | 2 | .014 | .038 | -.024 | -44 | 48 | -9 |
| 1 | 10 | 22 | 3 | .011 | .011 | .000 | 8 | 63 | 45 |
| 2 | 10 | 18 | 1 | .013 | .012 | .001 | 13 | -39 | -46 |
| 2 | 10 | 18 | 2 | .010 | .012 | -.002 | 13 | -46 | 31 |
| 2 | 10 | 18 | 3 | .007 | .014 | -.007 | -10 | -51 | -13 |
| 2 | 12 | 16 | 1 | .008 | .010 | -.002 | -12 | -40 | -31 |
| 2 | 12 | 16 | 2 | .011 | .014 | -.003 | 11 | -41 | 18 |
| 2 | 12 | 16 | 3 | .009 | .009 | .000 | 17 | -25 | -54 |
| 2 | 11 | 17 | 1 | .012 | .013 | -.001 | 6 | -31 | -55 |
| 2 | 11 | 17 | 2 | .009 | .012 | -.003 | -10 | -64 | -54 |
| 2 | 11 | 17 | 3 | .012 | .015 | -.003 | -12 | -52 | -74 |
| 3 | 12 | 22 | 1 | .063 | .016 | .047 | 99 | -98 | 58 |
| 3 | 12 | 22 | 2 | .019 | .017 | .002 | 46 | 66 | 37 |
| 3 | 12 | 22 | 3 | .026 | .017 | .009 | 38 | 62 | 62 |
| 3 | 10 | 24 | 1 | .021 | .014 | .007 | 38 | 118 | 36 |
| 3 | 10 | 24 | 2 | .023 | .015 | .008 | 39 | 96 | 52 |
| 3 | 10 | 24 | 3 | .028 | .016 | .012 | 36 | 72 | 61 |
| 3 | 11 | 23 | 1 | .029 | .011 | .018 | 14 | 72 | 42 |
| 3 | 11 | 23 | 2 | .026 | .012 | .014 | 25 | 89 | 62 |
| 3 | 11 | 23 | 3 | .022 | .012 | .010 | 27 | 82 | 65 |
| 4 | 12 | 22 | 1 | .012 | .021 | -.009 | 23 | 44 | -36 |
| 4 | 12 | 22 | 2 | .010 | .009 | .001 | 15 | 30 | 51 |
| 4 | 12 | 22 | 3 | .012 | .010 | .002 | 8 | 41 | 47 |
| 4 | 10 | 24 | 1 | .011 | .015 | -.004 | 5 | 26 | 29 |
| 4 | 10 | 24 | 2 | .013 | .011 | .002 | 6 | 61 | 37 |
| 4 | 10 | 24 | 3 | .011 | .017 | -.006 | -37 | 38 | 22 |
| 4 | 11 | 23 | 1 | .016 | .015 | .001 | 5 | 62 | 57 |
| 4 | 11 | 23 | 2 | .020 | .012 | .008 | 7 | 42 | 48 |
| 4 | 11 | 23 | 3 | .020 | .016 | .004 | 7 | 55 | 68 |

Table 8. Baseline and end-of-treatment voice samples randomly paired and spliced together to form a master listening tape for each of the subjects. The jitter difference values (col 7) were obtained by subtracting the end-of-treatment values from the end-of-baseline values for each of the paired samples of /a/.

direction consistent with the change in each of the subject's jitter levels. This conclusion is based on the results presented in Table 8. In terms of direction (sign agreement) judges one, two and three were able to rate all of the subjects' voices in the expected direction 75, 67, and 75% of the time, respectively. It is also apparent from this table that the judges were able to rate subject three's voice more consistently in the expected direction when compared to their ratings of the other subjects. This suggests that based on a purely qualitative judgement they may have been better able to rate subject three's voice in a way that relates to the fact that her acoustic data showed the most change in jitter with treatment.

Other studies have found significant correlations between jitter and perceptual ratings (Kane & Wellen, 1985; Askenfelt & Hammarberg, 1986). For example, Kane and Wellen (1985) measured jitter in the sustained /a/ productions of 10 children with vocal nodules. They compared these jitter measures to clinical ratings of severity of vocal disorder using the Wilson Voice Profile system (1971). These authors found that jitter magnitude correlated positively with the voice profile ratings. Askenfelt and Hammarberg (1986) measured jitter levels in the voiced segments of a reading passage spoken aloud by 41 patients with voice disorders, and compared them to perceptual ratings of voice made by speech-language pathologists for the same patients before and after therapy. The jitter values for the voiced segments of the passage were correlated with perceptual measures based on listeners' hearing of the entire reading passage. More significant correlations may have been reported for these studies than for the present investigation for several reasons. The perceptual ratings in the studies of Kane and Wellen (1985) and Askenfelt and Hammarberg (1986) were not based on the same context as the jitter measurements. Jitter was measured in sustained vowels, but perceptual ratings were based on running speech samples. This may have provided listeners with more cues to rate the voice samples. In addition, measurements of jitter may vary as a function of the speech sample analyzed. Fitch (1990) reported the consistency of the acoustic measurement of fundamental frequency and frequency perturbation using the Visi-pitch for samples of recorded speech including a reading passage, isolated vowels, and spontaneous speech. Results of this study indicated that the fundamental frequency of vowels was higher than the fundamental frequency of reading and spontaneous speech. Fundamental frequency is known to influence jitter values, so the jitter magnitudes may be different in different contexts. With respect to the present study, the relative changes that occurred in jitter with treatment were small and did not always change in a consistent direction with the change in nodules or the other acoustic variables. The samples of sustained vowels in the present study may have been more difficult to rate than the samples discussed above. In the last analysis, however, it is difficult to compare these data because Kane and Wellen (1985) and Askenfelt and Hammarberg (1986) used different perturbation formulae and acoustical analysis methods in calculating their values.

Correlations in the present research were influenced by both the relative size and direction of the jitter magnitudes and the direction and relative size of the perceptual difference ratings for each of the subjects. Although the judges were asked to rate the vowel pairs relative to one another, they were not asked specifically to attend to the magnitude of change in the data. They may have rated an end-of-treatment voice as better on the VAS, but the scale value of this difference (mm) may not have correlated with the same relative difference in the size of the jitter difference (in ms).

It is also possible that a combination of the acoustic variables is influencing the perceptual judgements. Hillenbrand (1988) measured the perception of aperiodicities in synthetically generated voice signals. He also found strong positive correlations between trained listeners' perceived ratings of roughness and pitch or amplitude variations. However, unlike the studies previously mentioned, Hillenbrand found that the perception associated with jitter was different from that associated with shimmer, and a strong relationship was found between the perception of breathiness and additive noise (SNR). How these acoustic variables (jitter, shimmer, $F_0$ and SNR) combine to influence listeners' judgements of the change or lack thereof in a subject's voice before and after treatment remains to be determined.

Interpretation of these results also must be tempered by the recognition that these findings have not been replicated, the sampled number of data points was small and the millimetre scale was arbitrarily applied to the VAS to determine correlations for the purpose of this discussion. Future studies attempting to correlate jitter with changes in perceived dysphonia might use an equal appearing interval scale or direct magnitude estimation on a larger sample and attempt to determine how these variables combine to influence listener judgements of the voice.

Research question number five concerned the question of what variable or combination of acoustic variables contributes to the perception of change in voice quality. This question was beyond the scope of this study. A multiple regression prediction equation could not be used to determine the relative contributions of each of the four acoustical variables (jitter, shimmer, $F_0$ and SNR) to the perception of change in voice quality over time due to the limited number of data points. This question remains an important one for future research.

In summary, results from the present study suggest that measures of jitter, shimmer, and SNR in the midportions of sustained productions of /a/ did not appear to change significantly or in a consistent direction with voice treatment. Fundamental frequency values showed more absolute change with treatment and tended to change in a manner more consistent with the change in each subject's vocal nodules. However due to the small sample size, the generalizability of these results is limited. It would be worthwhile for more clinicians to track changes in jitter, shimmer, $F_0$ and SNR associated with voice therapy as no firm conclusions can be made from the results of this study. A combination of jitter, shimmer, $F_0$ and SNR values may provide a better measurement of change in subjects with voice disorders. Perceptual judgement ratings of differences in dysphonia with treatment performed by speech-language pathologists did appear to correspond to changes in the status of each subject's vocal nodules, but did not appear to correlate well with changes in average jitter magnitudes. In the next section, limitations and suggestions for future research will be discussed relative to the research questions originally posed.

## Limitations and suggestions for future research

It is possible that the results obtained for the acoustical variables may have been different if subjects had consisted of a more homogeneous group, better controls for variables known to influence jitter and shimmer values had been used and different sampling procedures had been employed.

The present research attempted to control for voice onset/offset characteristics, time of day, practise effects, and any effects that might be observed as a result of the menstrual cycle as these are factors known to influence jitter levels. Within-session variability of jitter magnitudes is another factor to consider. Higgins and Saxman (1989) suggest that multiple sampling sessions may be required in order to obtain a more representative evaluation of a speaker's typical perturbation value. Linville and Korbacic (1987) suggest that intraspeaker variability on jitter measures in elderly women's sustained vowel productions can be significant. Linville and Korbacic (1987) suggest that as many as 15 trials may be required in order to obtain a representative sample. Although the subjects of the present research were not elderly, this information may be important in determining the minimum number of voice samples required to obtain a valid measure of jitter in other studies of pathologic voices. Future studies may wish to examine the within-session stability of jitter values in speakers with laryngeal pathologies.

Jitter values may also be sensitive to factors such as an individual subject's heartbeat (Orlikoff & Baken 1989; Orlikoff, 1990), psychological stress (Inbar & Eden, 1976; Eden & Inbar, 1978) and as well, colds or allergies. It is possible that the recording conditions in this study may have introduced some physical and psychological stress for some or all of the subjects. Subjects had to climb four flights of stairs to reach the room in which the recordings took place. It is possible that if subjects were not conditioned to climb the stairs at the beginning of the study their vocal jitter levels may have been elevated and more variable during early baseline recordings, and less so toward the end of baseline and during treatment when subjects were physically conditioned to climb the stairs. Individual stress associated with task performance, home or work experiences may also have influenced jitter. Future studies designed to measure jitter values may wish to examine the effects of physical and emotional stress on these measures.

Jitter values are also influenced by changes in fundamental frequency. Although the present study attempted to control for the influence of fundamental frequency by using a method to elicit the subject's habitual pitch level during data recordings across the entire experimental period, future studies might employ a more accurate tone pitch matching procedure to control for fundamental frequency effects on jitter. Because shimmer values are known to covary with jitter, the same considerations applied in the study of jitter must be applied to research designed to measure shimmer magnitudes.

Future investigations may also wish to explore changes in jitter and shimmer associated with different vowels and different voice disorders with treatment. These changes could be examined in the context of different voice samples which might include voice onset or offset characteristics, or both, in sustained vowels and connected speech. There also may be more demonstrable changes in the acoustical variables (as subject three's data suggest) associated with treatment depending on the severity of the vocal pathology at the outset of treatment. Perhaps the subject's nodules have to show a relatively large reduction in size before a significant difference is detected by analysis of the acoustic variables. Future studies may wish to explore the relationship between the amount of change in a subject's nodules or other pathology and the change in the acoustic variables more precisely.

The present study employed an A-B single-subject design to allow for pre-treatment tracking of the acoustic variables and documentation of their change across a period of treatment. The use of this single-subject design allows one to observe variability within subjects and to replicate treatment effects only across subjects. If stable baselines could have been obtained for each or any of the acoustic variables it may have been possible to replicate a treatment effect across subjects. The intraspeaker variation associated with a subject's ability to exert control over her perturbation measures, as well as the other previously mentioned speaker and measurement variables should encourage more clinicians to use single-subject designs to study these variables. If a treatment withdrawal component or better yet a withdrawal and second treatment period were added to the design a stronger case could be made for any treatment effects associated with any changes in the acoustic variables.

All listeners heard a difference in the severity of dysphonia between paired samples of sustained /a/ obtained during the last week of baseline and the last week of treatment. However, as mentioned previously, listeners commented on their general difficulty in rating subjects' voices. They suggested that perhaps if they had had a longer sample to judge with perhaps more cues (voice onset/offset, pitch breaks) they may have been better able to judge the voices. Future investigations may wish to explore the influence of varying the length and type of sample subjected to acoustical and perceptual analyses on the accuracy and reliability of listeners' perceptual ratings.

## Conclusion

In conclusion, results from the present study suggest that jitter, shimmer and SNR levels measured in the sustained /a/ productions of four adult females with vocal nodules before and during voice therapy did not show much absolute change or change in a consistent direction with treatment. The relative change in jitter, shimmer and SNR magnitudes did not appear to vary in a consistent manner with changes in the size of each subject's vocal nodules. In contrast to the results obtained for jitter, shimmer and SNR, fundamental frequency did demonstrate noticeable change with voice therapy. As well, changes in the direction of fundamental frequency appeared to be more consistent with the change in the size of the nodules for each of the subjects at the end of treatment. Fundamental frequency may provide a better index of change in the voice with voice therapy.

Although judges were able to rate the majority of voices in the expected direction at the end of treatment, jitter levels measured in subjects' sustained /a/ productions at the end of baseline and the end of treatment did not correlate well with the same perceptual ratings of relative dysphonia. The correlations between jitter difference values and perceived dysphonia were positive, but only significant for one judge's rating of one subject ($p < 0.05$).

Interpretation of these results must be tempered by the recognition that these findings have not been replicated, the sampled number of data points was small and the millimetre scale was arbitrarily applied to the VAS to convert the perceptual ratings into interval data for correlations for the purpose of discussion. Future studies attempting to correlate jitter with changes in perceived dysphonia might use other forms of perceptual scaling on a larger sample. It also remains to be determined as to how these acoustic variables (jitter, shimmer, $F_0$ and SNR) combined to influence listeners' judgements of the overall improvement or lack thereof in a subject's voice. Future investigations may wish to focus on which variable or combination of acoustic variables most influences our perceptual judgements of dysphonia.

# REFERENCES

American Speech and Hearing Association (1978). Guidelines for manual pure-tone audiometry. Asha, 20, 297-301.

Aronson, A.E. (1980). Clinical voice disorders. New York: Thieme-Stratton.

Askenfelt, A.G., & Hammarberg, B. (1986). Speech waveform perturbation analysis: a perceptual-acoustical comparison of seven measures. Journal of Speech and Hearing Research, 29, 50-64.

Baken, R.J. (1987). Clinical measurement of speech and voice. Boston: MA, College-Hill Press.

Beckett, R.L. (1969). Pitch perturbation as a function of subjective vocal constriction. Folia Phoniatrica, 21, 416-425.

Benjamin, B.J. (1981). Frequency variability in the aged voice. Journal of Gerontology, 36, 722-726.

Boone, D.R. (1971). The voice and voice therapy. (2nd ed.)., Englewood Cliffs, NJ: Prentice-Hall.

Boone, D.R. (1983) The voice and voice therapy. (3rd ed.)., Englewood Cliffs, NJ: Prentice-Hall.

Bond, A. & Lader, M. (1974). The use of analogue scales in rating subjective feelings. British Journal of Medical Psychology, 47, 211-218.

Brown, W.S. & Hollien, H. (1981). Effect of menstruation on fundamental frequency of female voices. In Transcripts of the tenth symposium on the care of the professional voice. Ed. V. Lawrence. New York: The Voice Foundation, 94-101.

Brown, W.S. & Hollien, H. (1982). Effects of menstruation on the singing voice. In Transcripts of the tenth symposium on the care of the professional voice. Ed. V. Lawrence. New York: The Voice Foundation, 140-147.

Brown, W.S. & Rothman, H. (1985). The menstrual cycle and its relationship to the highly stressed female voice. In Transcripts of the fourteenth symposium on the care of the professional voice. Ed. V. Lawrence. New York: The Voice Foundation, 114-123.

Bruning, J.L., & Kintz, B.L. (1977). Computational handbook of statistics. (2nd ed.)., Dallas, Tex.: Scott, Foresman & Company.

Coleman, R.F. (1969). Effect of median frequency levels upon the roughness of jittered stimuli. Journal of Speech and Hearing Research, 12, 330-336.

Crystal, T.H., & Jackson, T.L. (1970). Extracting and processing vocal pitch for laryngeal disorder detection. Journal of the Acoustical Society of America, 48, 118.

Davis, S.B. (1976). Computer evaluation of laryngeal pathology based on inverse filtering of speech. SCRL Monograph No. 13. Santa Barbara, California: Speech Communications Laboratory.

Davis, S.B. (1979). Acoustic characteristics of normal and pathological voices. In Lass, N.J. (Ed.) Speech and Language: Advances in Basic Research and Practise, vol. 1. New York Academic Press, 271-335.

Davis, S.B. (1981). Acoustical characteristics of normal and pathological voices. ASHA Rep., 11, 97-115.

Deal, R.Z. & Emanuel, F.W. (1978). Some waveform and special features of vowel roughness. Journal of Speech and Hearing Research, 21, 250-267.

Deem, J.F., Manning, W.H., Knack, J.V., & Matesich, J.S. (1989). The automatic extraction of pitch perturbation using microcomputers: some methodological considerations. Journal of Speech and Hearing Research, 32, 689-697..

Doherty, E.T. & Shipp, T. (1988). Tape recorder effects on jitter and shimmer extraction. Journal of Speech and Hearing Research, 31, 485-490.

Eskenazi, D.G., Childers, D.G., & Hicks, D.M. (1990). Acoustic correlates of vocal quality. Journal of Speech and Hearing Research, 33, 298-306.

Feijoo, S. & Hernandez, C. (1990). Short-term stability measures for the evaluation of vocal quality. Journal of Speech and Hearing Research, 33, 324-334.

Fitch, J.L. (1990). Consistency of fundamental frequency and perturbation in repeated phonations of sustained vowels, reading, and connected speech. Journal of Speech and Hearing Disorders, 55, 360-363.

Fairbanks, G. (1960). Voice and articulation drillbook. New York: Harper and Brothers.

Frable, M. (1962). Hoarseness, a symptom of premenstrual tension. Archives of Otolaryngology, 75, 80-82.

Gilbert, H.R., & Weismer, G.G. (1974). The effects of smoking on the speaking fundamental frequency of adult women. Journal of Psycholinguistic Research, 3, 225-231.

Guyatt, G.H., Berman, L.B., Townsend, M., & Taylor, D.W. (1985). Should study subjects see their previous responses? Journal of Chronic Diseases, 38, 1003-1007.

Haji, T., Horiguchi, S., Baer, T., & Gould, W.J. (1986). Frequency and amplitude perturbation analysis of electroglottograph during sustained phonation. Journal of the Acoustical Society of America, 80, 58-62.

Hartmann, E., & von Cramon, D. (1984). Acoustic measurement of voice quality in central dysphonia. Journal of Communication Disorders, 17, 425-440.

Hecker, M.H.L., & Kreul, E.J. (1971). Descriptions of the speech of patients with cancer of the vocal folds. Part I: Measures of fundamental frequency. Journal of the Acoustical Society of America, 49, 1275-1282.

Heiberger, V.L., & Horii, Y. (1982). Jitter and shimmer in sustained phonation. In N.J. Lass (Ed.). Speech and language: Advances in basic research and practice (pp. 299-332). New York: Academic Press.

Hersen, M. & Barlow, D.H. (1976). Single case experimental designs: Strategies for studying behavior change (2nd ed.). New York: Pergamon Press.

Hess, W.J. (1982). Algorithms and devices for pitch determination of speech signals. In Automatic speech analysis and recognition. Ed. J.R. Haton Dordrecht: D. Reidel Publishing Co., 49-67.

Hess, W.J. (1983). Pitch determination of speech signals: Algorithms and devices. New York: Springer Verlag.

Higgins, M.B. & Saxman, J.H. (1989). Variations in vocal frequency perturbation across the menstrual cycle. Journal of Voice, 3(3), 233-243.

Higgins, M.B. & Saxman, J.H. (1989). A comparison of intrasubject variation across sessions of three vocal frequency perturbation indices. Journal of the acoustical society of America, 86 (3), 911-916.

Hillenbrand, J. (1987). A methodological study of perturbation and additive voice in synthetically generated voice signals. Journal of Speech and Hearing Research, 30, 448-461.

Hillenbrand, J. (1988). Perception of aperiodicities in synthetically generated voices. Journal of the Acoustical Society of America, 83(6), 2361-2371.

Hirano, M., Hibi, S., Yoshida, H., Hirade, Y., Kasuya, H. & Kikuchi, Y. (1988). Acoustic analysis of pathological voice: some results of clinical application. Acta Otolaryngology (Stockholm), 105, 432-438.

Hollien, H., Michel, J. & Doherty, E.T. (1973). A method for analyzing vocal jitter in sustained phonation. Journal of Phonetics, 7, 85-91.

Holmes, J.N. (1962). The effect of simulating natural larynx behavior on the quality of synthetic speech. Speech Communications Seminar, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden.

Horii, Y. (1975). Some statistical characteristics of voice fundamental frequency. Journal of Speech and Hearing Research, 18, 192-201.

Horii, Y. (1979). Fundamental frequency perturbation observed in sustained phonation. Journal of Speech and Hearing Research, 22, 5-19.

Horii, Y. (1980). Vocal shimmer in sustained phonation. Journal of Speech and Hearing Research, 23, 202-209.

Horii, Y. (1982). Jitter and shimmer differences among sustained vowel phonations. Journal of Speech and Hearing Research, 25, 12-14.

Huck, S.W., Cormier, W.H. & Bounds, W.G., Jr. (1974). Reading Statistics and Research. New York: Harper & Row.

Inbar, G.F. & Eden, G. (1976). Psychological stress evaluators: EMG correlation with voice tremor. Biological Cybernetics, 24, 165-167.

Isshiki, N., Yanagihara, N. & Morimoto, M. (1966). Approach to the objective diagnosis of hoarseness. Folia Phoniatrica, 22, 393-400.

Iwata, S. & von Leden, H. (1970). Pitch perturbation in normal and pathologic voices. Folia Phoniatrica, 22, 413-424.

Jacob, L. (1968). A normative study of laryngeal jitter. Unpublished masters thesis, University of Kansas.

Johnson, K.W. & Michel, J.F. (1969). The effect of selected vowels on laryngeal jitter. Asha, 11, 96.

Kahane, J.C. (1981). Anatomic and physiologic changes in the aging peripheral speech mechanism. In D.S. Beasley & G.A. Davis (Eds.), Aging communication processes and disorders (pp. 21-62). New York, NY: Grune & Stratton, Inc.

Kane, M. & Wellen, C.J. (1985). Acoustical measurements and clinical judgements of vocal quality in children with vocal nodules. Folia Phoniatrica, 37, 53-57.

Karnell, M.P. (1991). Laryngeal perturbation analysis: Minimum length of analysis window. Journal of Speech and Hearing Research, 34,(3) 544-548..

Karnell, M.P., Scherer, R.S., & Fischer, L. (1991) Comparison of acoustic voice perturbation measures among three independent voice laboratories. Journal of Speech and Hearing Research, 34,(4) 781-790.

Kirk, R.E. (1982). Experimental design: procedures for the behavioural sciences. Monterey: Brooks/Cole Publishing.

Kitajima, K., Tanabe, M., Isshiki, N. (1975). Pitch perturbation in normal and pathologic voice. Studia Phonologica, IX, 25-32.

Koike, Y. (1973). Application of some acoustic measures for the evaluation of laryngeal dysfunction. Studia Phonologica VII, 17-23.

Liberman, P. (1961). Perturbations in vocal pitch. Journal of the Acoustical Society of America, 33, 591-603.

Lieberman, P. (1963). Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. Journal of the Acoustical Society of America, 35, 344-353.

Linville, S.E., & Korabic, E.W. (1987). Fundamental frequency stability characteristics of elderly women's voices. Journal of the Acoustical Society of America, 81, (4), 1196-1199..

Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A. & Zoghaib, C. (1989). The Communicative Effectiveness Index: development and psychometric evaluation of a functional communication measure for adult aphasia. Journal of Speech Disorders, 54, 113-124.

Ludlow, C.L., Bassich, C.J., Connor, N.P., Coulter, D.P., & Lee, Y.J. (1987). The validity of using phonatory jitter and shimmer to detect laryngeal pathology. In T. Baer, C. Sasaki, & K. Harris (Eds.). Laryngeal Function in Phonation and Respiration. Boston: Little, Brown & Company, 492-508.

McGlone, R.E. & Hollien, H. (1963). Vocal pitch characteristics of aged women. Journal of Speech and Hearing Research, 6, 164-170.

McReynolds, L.V., & Kerans, K.P. (1983). Single-Subject Experimental Designs in Communicative Disorders. Baltimore: University Park Press.

Milenkovic, P. (1987). Least mean square measures of voice perturbation. Journal of Speech and Hearing Research, 30, 529-538.

Milenkovic, P.(1989). Speech analysis systems for personal computers. Miniseminar presented at the American Speech and Hearing Association Convention, St. Louis, Missouri.

Montgomery, A.A. (1967). Laryngeal frequency under selected phonatory conditions. Unpublished doctoral dissertation, Purdue Univ., Lafayette, IN.

Moore, P. & Thompson, C.L. (1965). Comments on physiology of hoarseness. Archives of Otolaryngology, 81, 97-102.

Moore, P. & von Leden, H. (1958). Dynamic variations in the vibratory pattern in the normal larynx. Folia Phoniatrica, 10, 205-238.

Moser, M. & Kittel, G. (1977). Rechnergestitzte Tonhohenbestimmung. Folia Phoniatrica, 29, 119-126.

Murry, T. & Doherty, E.T. (1980). Selected acoustic characteristics of pathologic and normal speakers. Journal of Speech and Hearing Research, 23, 361-369.

Murry, T. & Large, J. (1979). Frequency perturbation in singers. In Transcripts of the Seventh Symposium on the Care of the Professional Voice. Ed. Lawrence, V. New York: The Voice Foundation, Vol. 1, 36-39.

Orlikoff, R.F. (1990). The relationship of age and cardiovascular health to certain acoustic characteristics of male voices. Journal of Speech and Hearing Research, 33, 450-457.

Orlikoff, R.F. (1990). Vowel amplitude variation associated with the heart cycle. Journal of Speech and Hearing Research, 33, 450-457.

Orlikoff, R.F., & Baken, R.J. (1990). Consideration of the relationship between the fundamental frequency of phonation and vocal jitter. Folia Phoniatrica, 42, 31-40.

Orlikoff, R.F., & Baken, R.J. (1989). The effect of the heartbeat on vocal fundamental frequency perturbation. Journal of Speech and Hearing Research, 32, 576-582..

Ottenbacher, K.J. (1986). Evaluating clinical change: Strategies for occupational and Physical Therapists. Baltimore: Williams and Wilkins.

Pruszewicz, A., Obrebowski, A., Swidzinski, P., Demenko, G., Wika, T. & Wojciechowska, A. (1991). Usefulness of acoustic studies on the differential diagnostics of organic and functional dysphonia. Acta Otolaryngologica (Stockholm), 111, 414-419.

Ramig, L.A. (1980). Acoustic characteristics of speech and selected measures of body physiology. Unpublished doctoral dissertation, Purdue University.

Ramig, L.A. & Ringel, R.L. (1983). Effect of physiological aging on selected acoustic characteristics of voice. Journal of Speech and Hearing Research, 26, 22-30.

Ramig, L.A., Scherer, R.C., Titze, I.R., & Ringel, S.P. (1988). Acoustic analysis of voices of patients with neurologic disease: rationale and preliminary data. Annals of Otology, Rhinology and Laryngology, 97, 164-172.

Ramig, L.A., Scherer, R.C., Klasner, E.R., Titze, I.R., & Horii, Y. (1990). Acoustic analysis of voice in amyotrophic lateral sclerosis: A longitudinal case study. Journal of Speech and Hearing Disorders, 55, 2-14.

Read, C., Buder, E.H., & Kent, R.D. (1990). Speech analysis systems: A survey. Journal of Speech and Hearing Research, 33, 363-374.

Rozsypal, A. & Millar, B. (1979). Perception of jitter and shimmer in synthetic vowels. Journal of Phonetics, 7, 343-355.

Silverman, E. & Zimmer, C. (1978). Effect of the menstrual cycle on voice quality. Archives of Otolaryngology, 104, 7-10.

Sorensen, D. & Horii, Y. (1983). Frequency and amplitude perturbation in the voices of female speakers. Journal of Communication Disorders, 16, 57-61.

Takahashi, H. & Koike, Y. (1975). Some perceptual dimensions and acoustical correlates of pathologic voices. Acta Otolaryngologica, Suppl. 338, 1-24.

Titze, I.R., Horii, Y. & Scherer, R.C. (1987). Some technical considerations in voice perturbation measurements. Journal of Speech and Hearing Research, 30, 252-260.

Wendall, R.W. (1963). Laryngeal analog synthesis of harsh voice quality. Folia Phoniatrica, 15, 241-250.

Wilcox, K.A. & Horii, Y. (1980). Age and changes in vocal jitter. Journal of Gerontology, 35, 194-198.

Wilde, M.D., Martens, W.L., Hillenbrand, J. & Jones, D.R. (1986). Externalization mediates changes in the perceived roughness of sound signals with jittered fundamental frequencies, In Proceedings of the 1986 International Computer Music Conference, The Hague.

Wilson, F.B. (1971). The voice-disordered child; a descriptive approach. Language Speech and Hearing Services in Schools, 1, 14-22.

Wilson, F. & Purvis, J. (1980). A study of selected vocal behaviors during the menstrual cycle of trained singers. Journal of Research in Singing. 3, 15-24.

Wolfe, V., Cornell, R. & Palmer, C. (1991). Acoustic correlates of pathologic voice types. Journal of Speech and Hearing Research, 34,(3) 509-516.

Yanagihara, N. (1967). Significance of harmonic changes and noise components in hoarseness. Journal of Speech and Hearing Research, 10, 531-541.

Zajac, D. J., & Linville, R.N. (1989). Voice perturbations of children with perceived nasality and hoarseness. Cleft Palate Journal, 26, (3), 226-232.

Zemlin, W.R. (1962). A comparison of the periodic function of vocal fold vibration in a multiple sclerosis and a normal population. Unpublished doctoral dissertation, University of Minnesota.

Zyski, B.J., Bull, G.L., McDonald, W.E. & Johns, M.E. (1984). Perturbation analysis of normal and pathologic larynges. Folia Phoniatrica, 36, 190-198.

# APPENDIX A

## SUBJECT CONSENT FORM
## FOR INITIAL CONTACT BY THE PRIMARY INVESTIGATOR

To whom it may concern,

In this package you will find information about a research project involving subjects with vocal nodules and a subject consent form. If you think you are interested in participating in this research but would like more information, please sign below. Your signature grants the primary investigator permission to contact you by telephone, and at that time provide you with more information and answer any questions you may have regarding the study.

Subject's signature

Date

Telephone Number

Subject's printed name

Witness

# APPENDIX B

## TELEPHONE SCREENING FORM

Name ______________________ Date of Screening ______________

Address __________________________ Telephone ______________

DOB ____________ Age _____ Occupation ____________________

Referred by ____________________

Date of Flexible Fiberoptic Video Nasendoscopy (FFVN)
confirming presence of Vocal Nodules ___________________________

==============================================================

1) How long have you noticed a problem with your voice?

2) Have you ever had any voice or sp/lang therapy in the past?

3) Are you a smoker?

4) Have you ever had any respiratory illness or chronic allergies?

5) Are you taking any medications (includes antidepressants, antipsychotics, antihistamines, birth control pills)? If so, what are they?

6) Do you use any street drugs? If so, how often?

7) Are you willing to commit to a twelve week pre-treatment and treatment period?

8) Do you have any questions re: the study?

# APPENDIX C

**Vocal Nodules Treatment Study**
**Department of Speech Pathology and Audiology**
**Faculty of Rehabilitation Medicine**
**University of Alberta, Edmonton**
**Principal investigator: Katherine Ward, B.Sc.**
**(403) 437-7904 (Res)**
**(403) 492-5990 (Ofc)**
**Advisor: Anne Putnam Rochet, Ph.D.**

## INFORMATION FOR POTENTIAL PARTICIPANTS

### PURPOSE

This document describes a research project concerned with changes in the voice associated with voice therapy for vocal nodules. The purpose of this research project is to determine the effectiveness of different measures for monitoring voice therapy by: 1) using computer analysis to follow the changes in the voices of patients with vocal nodules before and during therapy, and 2) having individual speech pathologists rate the severity of "hoarseness" in patients with nodules before and during therapy.

The results of this study should help clinicians provide better ways of monitoring changes in patients' voices with therapy.

### Who is eligible to participate?

To be eligible as a subject, you must: 1) have vocal nodules associated with chronic hoarseness confirmed via flexible fiberoptic video nasendoscopy through the Glenrose Rehabilitation Hospital's Voice Clinic or U of A's Voice Clinic, 2) be female and between 18 and 35 years of age, 3) have abstained from smoking for at least two years, 4) have no history of respiratory illness, chronic allergies, and/or speech and hearing problems, and 5) have abstained from taking drugs (e.g., antipsychotics, antidepressants or antihistamines). In addition, you must pass a pure tone hearing screening test.

### Procedures for before, during and after treatment sessions

Your participation in the study will span twelve weeks during which you will be seen 3 times per week: This time period is divided into four weeks of pre-treatment monitoring, and eight weeks of voice treatment . Because this is a research project concerned with monitoring changes in the voice that occur with therapy, it is vital that each participant be followed for a length of time before therapy that is comparable to the minimum amount of time actually spent in therapy. The choice of 4 weeks as an experimental period is also influenced by the belief that the menstrual cycle may influence changes in the voice. Therefore, it is necessary to have each phase of the project encompass at least one complete menstrual cycle.

Information will be collected on your voice across the twelve weeks of the study by means of audio tape recordings. During the four week pre-treatment phase of the study you will be asked to come in 3 times per week for sessions that will last approximately 15-20 minutes.

During the eight week treatment phase of the study you will be asked to come in 3 times per week for sessions that will last 1 hour. The first 10 minutes of each session will be used to record your voice. During the remaining 50 minutes you will learn new ways of using your voice that counteract the kinds of vocal behaviors that are thought to cause vocal nodules.

As soon as possible after treatment ends, a second flexible fiberoptic video nasendoscopy exam will be scheduled at the Glenrose Voice Clinic or University of Alberta to document any change in the appearance of your vocal nodules.

Appointments can be scheduled at your convenience, either during the day or in the evening. A sum of **$25.00** will be paid to participants to cover parking expenses associated with the pre-treatment phase of the project.

The research will be conducted in the facilities of the Faculty of Rehabilitation Medicine on the University of Alberta campus in Edmonton. The execution of the research and the conduct of its investigator are governed by the University of Alberta's policy related to ethics in human research. In accordance with those regulations, your welfare and dignity will be protected during the project, and your anonymity will be guaranteed in any professional publications or lectures derived from it. The data derived from your participation `will be coded and kept in a secure place accessible only to the primary investigator.

There are no liabilities to you as a participant in this research, although it will require a considerable time commitment on your part. The flexible fiberoptic video nasendoscopy exam, which you experienced as part of the diagnostic procedures at the Glenrose Rehabilitation Hospital Voice Clinic or the U of A Clinic, is a necessary follow-up procedure that will be performed again at one of the Clinics as near the end of your treatment as possible. It will be performed using a topical anaesthetic to make it as comfortable for you as possible.

Your inquiring about this project in no way obliges you to participate in it. If you do consent, you are free to withdraw your participation at any time, without ill-will or jeopardy to your future health care at the Glenrose Rehabilitation Hospital, the University of Alberta or its Hospitals. Feel free to ask any questions about the research. Kate Ward is the principal investigator responsible for this project and its execution. Kate can be reached at 437-7904 (Res) or 492-5990 (Ofc).

Thank-you for your consideration of this information.

**Vocal Nodules Treatment Study**
**Katherine Ward, B.Sc., S-LP (C)**
**Department of Speech Pathology and Audiology**
**400 Garneau Professional Centre**
**University of Alberta**
**Edmonton, AB**
**T6G 0T2**

# APPENDIX D

**Vocal Nodules Treatment Study**
**Faculty of Rehabilitation Medicine**
**University of Alberta, Edmonton**

**Principal investigator: Katherine Ward, B.Sc.**
**Advisor: Anne Putnam Rochet, Ph.D.**

## SUBJECT'S CONSENT FORM

**Name:**_____________________________________

I have read the attached information regarding the vocal nodules treatment study to be conducted by Kate Ward. I understand the contents of the information form, the procedures involved in the project, the purpose of the research and the time commitment required. I have felt free to ask questions about the project, and they have been answered to my satisfaction.

I understand that:

1) An initial interview will confirm my eligibility for this treatment study. At that time I will be asked questions concerning my history of smoking, alcohol or other drug use, voice use habits, respiratory problems, chronic allergies, history of neurological disease and speech or hearing problems. This information will be recorded on my voice case history form. I will then undergo a brief hearing screening test which I must pass to be eligible to participate in the study.

2) The vocal nodules treatment project spans 12 weeks during which I will visit the project's Research Office on the University of Alberta's campus three times a week. The visits during the first four weeks preceding therapy will involve simply having my voice recorded as I perform various speaking tasks, and each visit will last approximately 15-20 minutes.

During the 8 weeks of treatment I will visit the Research Office three times per week for one-hour sessions. My voice will be recorded at the beginning of each session, and then I will receive approximately 50 minutes of voice therapy. The pre-and post-treatment recording periods are required to provide valid comparisons with the recordings during the treatment period.

3) I will be paid a sum of **$25.00** to cover parking expenses during the pre-treatment phase of the project.

4) The only invasive procedure associated with this research project is flexible fiberoptic video nasendoscopy. I have experienced this procedure as part of the diagnosis of my vocal nodules. In order to document the effectiveness of voice therapy I will be asked to consent to undergo this procedure once more at the Glenrose Rehabilitation Hospital's Voice Clinic or University of Alberta's Voice Clinic immediately following voice therapy. A topical anaesthetic will be applied to make this procedure as comfortable as possible.

5) Information derived from my participation in this research will help clinicians provide better ways of monitoring and providing efficient treatment for persons with voice disorders.

6) The execution of this project and the conduct of its personnel are governed by the University of Alberta's policy related to ethics in human research. In accordance with that policy, my welfare will be protected during the course of this project, and my anonymity will be assured in any professional publications or lectures derived from it. The data will be coded and kept in a secure place accessible only to the primary investigator.

I will attempt to stay with this project for the 8 to 12-week duration of pre-treatment and treatment measures. Nevertheless, I understand that I am free to withdraw my participation at any time without ill-will and without jeopardizing my future health care at the Glenrose Rehabilitation Hospital, the University of Alberta or its Hospitals. My signature below affirms my voluntary consent to participate and acknowledges my receipt of a copy of this consent form and an information form.

| | |
|---|---|
| Subject's signature | Date |
| Subject's printed name | Witness |
| | Principal Investigator's signature |

# APPENDIX E

## VOICE CASE HISTORY

Name ____________________ DOB _________ Age ___ Height _______

Address ________________________________ Telephone ____________

Referred by ____________________ Investigator _____________________

Date of Evaluation __________ Date of Evaluation at Voice Clinic ___________

==========================================================

**History of the Voice Problem** (adapted from D. Boone (1971))

1) Description of Problem:

2) Cause of Problem:

3) Onset of Problem:

4) Variability Through Day:

5) Voice Usage:

6) Known Abuse/Misuse:

**Health History:**

1) General Health:

2) Physical Problems related to Voice:

3) Illnesses and Allergies:

4) Accidents:

5) Surgery:

6) Medications (i.e., antidepressants, antihistamines, antipsychotics, birth control pills, stimulants, anticonvulsion drugs)

current:

in the last month:

7) Smoking/Alcohol:

8) History of Voice Problems:

9) Previous Voice Therapy:

**Hearing Evaluation:**

1) Pure-tone audiometric screening

Date ________________________

**Behavior Observations:**

Picture of Vocal Nodules (confirm via videofiberoptic nasendoscopy)

Date: ____________________________

# APPENDIX F

## VOICE TREATMENT INFORMATION FOR SUBJECTS

### SUGGESTIONS FOR IMPROVED VOCAL USE
### QUESTIONS AND ANSWERS REGARDING VOCAL NODULES

Kenneth W. Burk, Ph.D.
Wichita State University
Speech-Language Hearing Clinic

1. What are vocal nodules? Added layers of tissue on the vibrating edge of the vocal folds that vary in size from pinpoint to the size of a peppercorn. They develop as the body attempts to protect itself against abuse and overuse of the voice. They usually are on both vocal folds, and located a third of the way down from the front of the vocal folds.
2. How do they develop? Nodules develop through a person (a) continuously using a loud voice (whether speaking or singing), and (b) along with this, abusing the voice through shouting, yelling etc.
3. What is the primary cause? Because of what the person does with voice, the repeated hard beginning of voice (vocal attack), using a greater than normal loudness IS THE BASIS FOR NODULES DEVELOPING. This development can be enhanced by medical problems, e.g., laryngitis, edema (swelling), sinus, etc.
4. What are the symptoms? The major symptoms are that the person is having to work harder than normal to make voice (tension), and because of the presence of the nodules, is not able to get complete closure of the vocal folds (hence, there is breathiness). In addition, the person may have pitch breaks or uncertainty of pitch ( as during singing), a reduction in the upper part of the pitch range, as well as instances where the pitch level is heard lower than usual.
5. What is worked on directly? The symptom worked on directly is the loud abrupt vocal attack. No direct work is done on reducing the breathiness, changing pitch level. Along with changing the way voice is started (attack), there must be some reduction in the overall amount of talking or using the voice.

6. How long will it take? With good cooperation, an audible change in voice should be heard within 4-6 weeks. By 6 months, there should be an absence or significant reduction in the size of the nodules, as determined by medical examination.

7. How will I know the voice is improving? In order, the changes in voice are as follows: (a) reduction in pitch breaks and breaks in voicing, (b) awareness of less tension in making voice, and (c) an awareness of less breathiness in the voice.

## SUGGESTIONS FOR VOCAL HYGIENE

A) CONTROLLING BACKGROUND NOISE A basic concept to remember is that we reflexively increase our vocal loudness in the presence of a noisy environment. Rarely are we aware of it at the time.

a) A good example is that of an extended conversation in a car, especially when you are talking to a person in the back seat. When we are on the highway at the legal speed limit, with any of the following in operation-- heater, air conditioner, radio, we will double our vocal loudness. Only when we pull off the road are we aware that we have been shouting to one another.

b) The same holds true when you are competing with the television, radio, machine noise (fans, tools, etc.), or other background music.

c). A similar situation is that presented by trying to carry on an extended conversation over the noise of many in the group. Sometimes it calls for moving away to a more favorable place. Getting someone's attention in these situations may be best if non-verbal, e.g., hand gestures, tapping, etc.

B) LIMITING VOICE USE WHEN NECESSARY A good rule of thumb is that speaking in your normal voice at conversational volume promotes good vocal hygiene. Even using a slightly breathy voice when the opportunity is available -- and not costly personally or professionally, is one way of buying a vocal insurance policy against a time when you will have heavy demands on voice.

a) A use of voice that is loud and requires an abrupt attack (like a cheerleader's yell), will be costly to the normal vocal health. Excessive laughing or crying is also cautioned.

b) Whenever possible, avoid long conversations over poor connections that require you to use a louder than normal voice. The same thing might be said of talking at length with a person with a significant hearing loss.

c) While an attorney may not always have the opportunity, the use of amplification when talking with groups (or in court), may be a "voice saver" unless it is viewed as interfering with your professional style.

d) Instead of coughing and throat clearing (unless there is really something there to bring up), substitute a swallow or taking a sip of something to drink.

e) If you are a singer, bypass this activity when you have some active medical problem that affects the larynx. If fact, this holds true for the amount and type of talking that could lead to further irritation.

C) CREATING A GOOD VOCAL ENVIRONMENT It is well to remember that abuse to the vocal folds increases if the mucous coverings are allowed to dry out or to become further irritated.

a) Increase the humidity at home (and many do at the office as well), to 40-50% through the use of various cold-air humidifiers, plants, pans of water near hot air registers, etc.

b) If you do smoke, it is wise to significantly reduce or eliminate this -- especially important for persons who earn their livings with the voice. This includes exposure to cigarette smoke.

c) Limit exposure to aerosol sprays of any kind (perfume, hair spray, deodorant, bathroom or oven cleaners, room fresheners, disinfectant sprays, etc.)

d) Restrict intake of irritants in beverages containing alcohol or caffeine (colas, tea, coffee, etc.). All of these in excessive amounts and during times of heavy use of voice, tend to dry out the mucous membrane linings of the throat and larynx. Also, many of the medications to handle allergies (antihistamines), and sinus drainage, tend also to dry out these linings. It is rare that a performer will go on stage just having taken a heavy dose of antihistamines.

e) In general, one should restrict intake of "hot" drinks" -- at times of heavy use, since they tend to dry as well.

D) INCREASE LUBRICATION OF THE VOCAL FOLDS In the previous part of these suggestions, many were aimed at reducing factors that dry out membranes of the larynx and vocal folds. There are some ways of increasing the moisture level of the mouth and larynx. One of these ways is not to use cough drops with a menthol base -- since these function to dry out rather than to increase moisture.

a) When breathing through the mouth, place the tongue on the roof of the mouth behind the upper teeth to increase the humidity of the air being inhaled.

b) Cut fresh lemon into six (6) wedges and chew one wedge six times a day. Citric acid has a documented positive effect on moisture of the larynx.

c) Most professional voice users will drink 7-9 glasses of water a day. When performing they are likely to drink water at room temperature -- as opposed to extremely hot or cold fluids or drinks.

d) Chewing gum on occasion has a tendency to increase moisture.

e) Some report that SORBEE, a sugarless lozenge, stimulates an increased amount of moisture. (Available at health-food stores.)

## E) SOME FINAL LIFE AND VOCAL STYLE SUGGESTIONS

a) It is suggested that you keep track of acidity levels in the body and handle them as you may decide. At the level of reflux, it can constitute a source of irritation to the vocal mechanism.

b) Since the encasement of the voice is the body, it would be well to get some consistent and regular exercise. Competitive sports are not encouraged, but swimming is encouraged.

c) Be careful of overuse of common medications. Aspirin, for example, acts as an anticoagulant and may in some circumstances be a potential problem for a performer.

d) Finally, many heavy voice users reduce significantly the amount of starches in the diet, and emphasize fresh fruits and vegetables. Excessive consumption of milk-products is sometimes not encouraged.

### Vocal Use Practices

1. When singing or using the speaking voice in unusual ways (i.e., other than in quiet conversation) and in circumstances such as stage acting, public speaking, cheerleading, lecturing in class, or occupational uses such as sales work, learn to use your voice with as little effort and tension as is necessary. In other words, avoid hyperfunctional use of your voice. Less is more if we think of vocal longevity.

2. Avoid singing or speaking for excessive amounts of time. Keep in mind that voicing for long periods of time related directly to the physical conditioning of the vocal muscles and also to the innate vocal capacity of the individual to endure the wear and tear.

3. Avoid singing in a tessitura which is continually near the extremes of your own vocal range, and carefully pace the use of register extreme such as "pushing the chest voice into the upper range." Overuse here can be vocal suicide.

4. Before singing or using the voice in unusual ways, always do vocal warm-ups. Exercises should begin with general muscle stretching and massaging, etc., and vocal exercises should proceed from less strenuous to more strenuous usage. Loud volume and high pitch are the most strenuous uses of voice, so begin warm-ups by exercising the mid and low range first, softly. Gradually include louder and higher pitched vocal usage with easy production. Warm up for about 7-10 minutes; 5 minutes should be the absolute minimum. Different voices have different needs but every voice needs warm-ups.

5. Reduce general voice use before a performance. Reduce your mileage. Think conservation.

6. Avoid shouting, screaming, loud laughter, and heavy throat clearing.

7. Hold down coughing, and when sneezing, do it gently and non-vocally.

8. Avoid talking in noise, i.e., in loud restaurants, or Junior League with band, and watch talking over motor noise in cars, planes and busses. Keep your vocal mileage, again, as low as possible.

9. Avoid making strange vocal noises and sounds.

10. Do not sing with a sore throat.

11. If it feels bad, don't do it.

Common Signs of Significant Vocal Abuse

1. Throat is tender to the touch after vocal use.

2. Hoarse voice at the end of singing.

3. Need to clear the throat for a period of time.

4. Throat is very dry, has a noticeable "tickle" that is persistent (check out dehydration).

5. Inability to produce your highest notes at pianissimo volume.

6. Persistent hoarseness or inability to sing or speak with a clear voice after 24-48 hours of voice rest.

## VOICE CHART

| | SUN | MON | TUES | WED | THURS | FRI | SAT |
|---|---|---|---|---|---|---|---|
| **Drink 7-9 glasses of water/day** | | | | | | | |
| **Throat clearing** | | | | | | | |
| **Coughing** | | | | | | | |
| **Yelling/Cheering** | | | | | | | |
| **Loud Laughing** | | | | | | | |
| **Talking over noise** | | | | | | | |
| **Excessive talking** | | | | | | | |

| | SUN | MON | TUES | WED | THURS | FRI | SAT |
|---|---|---|---|---|---|---|---|
| **Drink 7-9 glasses of water/day** | | | | | | | |
| **Throat clearing** | | | | | | | |
| **Coughing** | | | | | | | |
| **Yelling/Cheering** | | | | | | | |
| **Loud Laughing** | | | | | | | |
| **Talking over noise** | | | | | | | |
| **Excessive talking** | | | | | | | |

# APPENDIX G

## INSTRUCTIONS FOR PERCEPTUAL RATINGS

You are being asked to make perceptual judgments about voice samples obtained from four subjects who received voice therapy for the treatment of vocal nodules. Voice samples of these subjects sustaining the vowel /a/ were recorded during a pre-treatment baseline and across the duration of voice therapy. Each voice sample that you will hear will be approximately 3 seconds in length. The mid three seconds of each vowel were chosen to avoid voice onset and offset characteristics. You may detect some abruptness at the beginning or end of the signals due to the tape splicing of the samples. These samples will be presented to you in pairs, one of which was taken from the final week of baseline and one from the final week of treatment. The order in which you hear the components of a pair will be randomized on the tape. Thus, sometimes you will hear the end-of-baseline sample first and then the end-of-treatment sample, and sometimes the end-of-treatment sample first and then the end-of-baseline sample. This is done to avoid contaminating your perceptual ratings with expectations about treatment outcome. Please just rate the components of a voice sample pair as objectively as possible according to the directions below.

A visual analogue scale has been provided for you to rate each of the paired samples on a continuum from normal to dysphonic voice. You will be listening for changes in pitch, pitch breaks, breathiness, hoarseness, diplophonia, and glottal fry. Listen to each pair of voice samples and then mark a "1" somewhere on the scale for the first /a/ of the pair and a "2" for the second /a/ of the pair to indicate where it falls on the scale relative to your judgment of the first /a/. I will stop the tape between pairs to allow you sufficient time to mark on the scale. If you require a replay of a vowel pair in order to make an accurate rating please let me know during the session. You are also asked to indicate on the same page which vocal quality (ies) influenced your perceptual rating the most for each vowel pair. A practice listening tape will be presented first to allow you time to become familiar and comfortable with the perceptual rating procedure.

Following a short break, the entire master listening tape for each subject will be replayed for you to establish intrajudge reliability for the perceptual ratings.